

UNIVERSIDAD ESAN



**Metodología de Implementación del Credit Scoring en una entidad
Financiera en el Segmento Pyme**

**Tesis presentada para cumplir con uno de los requisitos para la obtención
del grado académico de Maestro en Finanzas por:**

Andy Flores Bardales

Programa de la Maestría en Finanzas

Lima, 13 de febrero de 2019

Esta tesis

**Metodología de Implementación del Credit Scoring en una entidad
financiera en el Segmento Pyme**

ha sido aprobada.

.....

César Fuentes, Ph.D. (Jurado)

.....

Luis Chávez Bedoya, Ph.D. (Jurado)

.....

Alfredo Mendiola, Ph.D. (Asesor)

.....

Carlos Aguirre, Mg. (Asesor)

Universidad ESAN

2019

DEDICATORIA

A mi papito Dios, que está reconstruyendo mi vida, gracias por la vida, tu amor, misericordia y providencia. Concédeme un corazón agradecido y no dudar nunca de tu amor.

A mis padres Raúl y María, gracias por regalarme el don de la vida, su amor, cariño y apoyo. A mi hermano Paúl gracias por compartir conmigo muchos momentos felices y también dificultades, te quiero mucho. A mis hermanos de parte de mi papa Raúl, con afecto y cariño sincero.

A mi esposa Milagros y a mis hijos Juan Diego, Valeria, Santiago Joaquín, Paula Francesca y María Gracia Del Carmen, gracias por su amor, cariño y apoyo. Todos son un regalo de Dios y los llevo en el corazón.

A mis hermanos de comunidad de la Parroquia Sagrados Corazones de Jesús y María que siempre rezan por mí.

Andy Flores Bardales

ÍNDICE GENERAL

ÍNDICE DE FIGURAS.....	vii
ÍNDICE DE TABLAS	viii
CURRICULUM VITAE.....	x
CAPÍTULO I. INTRODUCCIÓN.....	iii
1.1. Introducción	iii
1.2. Objetivo.....	1
1.3. Objetivos específicos	1
CAPÍTULO II. CREDIT SCORING	2
2.1. Proceso Crediticio	2
2.2. Definición de Credit Scoring	5
2.3. Propósito del Credit Scoring	7
2.4. Aplicaciones del Credit Scoring	8
2.5. Ventajas y desventajas del Credit Scoring.....	9
2.6. Tipos de Credit Scoring	10
2.7. Acuerdos de Capital Basilea I, II, III y el Credit Scoring.....	11
CAPÍTULO III. PYMES.....	12
3.1. Definición de Pyme.....	12
3.2. Característica de la Pyme	13
3.3. Tipo de Créditos (Pyme).....	13
3.3.1 <i>Créditos a Pequeñas Empresas</i>	14
3.3.2 <i>Créditos a Microempresas</i>	14
3.4. Categorías de Clasificación Crediticia del Deudor de Créditos a Pequeñas Empresas, a Microempresas, de Consumo Revolvente y No Revolvente	14
3.5. Productos ofrecidos a las Pymes	15
3.6. Indicadores Pyme	16
CAPÍTULO IV. MODELOS Y METODOLOGÍAS DEL CREDIT SCORING	19
4.1. Modelos Paramétricos Multivariantes	19
4.2. Modelos No Paramétricos.....	26
4.3. Modelos Semi-Paramétricos (Hybrid Models)	28

4.4. Ventajas y desventajas de los Modelos Paramétricos, No Paramétricos y Semi Paramétricos	32
4.5. Elección de función de enlace.....	34
4.6. Análisis de la función de enlace.....	34
4.7. Curva logística	36
4.8. Suposiciones de la regresión logística	37
4.9. Estimación de máxima verosimilitud.....	38
4.10. Evaluación del modelo de regresión logístico binario	39
4.10.1 Evaluación general del modelo	39
4.10.2 Significancia estadística de los coeficientes de regresión individuales	41
4.10.3 Precisión predictiva y discriminación	42
4.11. Medición del buen desempeño de la regresión Logística	44
CAPÍTULO V. CREDIT SCORING EN PYMES.....	46
5.1. Aplicaciones del Credit Scoring	46
5.2. Selección de variables	58
CAPÍTULO VI. APLICACIÓN DEL CREDIT SCORING EN PYMES	64
6.1. Aplicación del Credit Scoring en el producto Capital de Trabajo Revolvente.....	64
6.1.1 Definición de la Población Objetivo.....	65
6.1.2 Limitaciones del Modelo.....	65
6.1.3 Variables del Credit Scoring revolvente:	66
6.1.4 Pruebas de coherencia.....	67
6.1.5 Definición de Incumplimiento.....	68
6.1.6 Ventana de Observación.....	69
6.1.7 Segmentación.....	70
6.1.8 Características de la muestra	70
6.2. Principales características que influyen en el modelo Credit Scoring	72
6.3. Credit Scoring Revolvente	80
6.4. Performance del Credit Scoring Pyme Revolvente.....	82
6.5. Tabla de clasificación muestra test Pyme Revolvente	82
6.6. Curva ROC del Credit Scoring Pyme Revolvente en la muestra test	83
6.7. Validación del Credit Scoring Revolvente en personas naturales y jurídicas.....	83

6.8. Credit Scoring con solo variables del Titular de la Pyme.....	86
6.9. Validación del Modelo.....	89
6.9.1 Indicadores de Bondad de ajuste del Credit Scoring Revolvente.....	89
6.9.2 Intervalos de confianza para el Credit Scoring Revolvente	89
CAPÍTULO VII. CONCLUSIONES	92
CAPÍTULO VIII. RECOMENDACIONES.....	99
ANEXOS	100
BIBLIOGRAFÍA	106

ÍNDICE DE FIGURAS

Figura II.1 Proceso de promoción, evaluación y concesión de un microcrédito	2
Figura IV.1 Créditos Directos Pyme.....	17
Figura IV.2 Calidad de la Cartera Pyme	17
Figura IV.3 Cobertura Pyme.....	18
Figura III.1 Función logística.....	37
Figura III.2 Ejemplo de la curva ROC	44
Figura III.3 Indicador Gini.....	45
Figura VI.1 Default vista 12 meses.....	68
Figura VI.2 Ventana de Observación del Credit Scoring	69
Figura VI.3 Tasa de Default por ventiles ratio de endeudamiento	73
Figura VI.4 Tasa de default por ventiles de número de empresas reportado.....	73
Figura VI.5 Tasa de default por ventiles de sobregiro.....	74
Figura VI.6 Tasa de default por ventiles evolución de deuda en el sistema	74
Figura VI.7 Tasa de default por ventiles pasivo entre patrimonio.....	75
Figura VI.8 Tasa de default por ventiles capital de trabajo neto	75
Figura VI.9 Tasa de default por ventiles de patrimonio	76
Figura VI.10 Tasa de default por ventiles saldo activo vencido.....	76
Figura VI.11 Tasa de default por ventiles de diferencia entre meses con buen comportamiento y mal comportamiento en el sistema financiero	77
Figura VI.12 Tasa de default por ventiles peor clasificación de riesgo en los últimos 24 reportes en el sistema financiero	77
Figura VI.13 Tasa de default por ventiles del pasivo	78
Figura VI.14 Variable zona geográfica.....	78
Figura VI.15 Tasa de default por ventiles de meses de constitución.....	79
Figura VI.16 Tasa de default por ventiles de antigüedad	79
Figura VI.17 Curva Roc Pyme Revolvente	83
Figura VI.18 Matriz de Correlaciones Modelo Revolvente vs Modelo Variables del Titular.....	88
Figura VI.19 Intervalo de confianza muestra train y test.....	90

ÍNDICE DE TABLAS

Tabla I.1 Ratio de morosidad del sistema financiero.....	iv
Tabla IV.1 Definición de Pyme	13
Tabla III.1 Ventajas y desventajas de los modelos paramétricos, no paramétricos y semi paramétricos	32
Tabla III.2 Tabla de clasificación de muestra.....	43
Tabla V.1 Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II.....	47
Tabla V.2 Modelo Scoring para el otorgamiento de crédito de las Pymes.....	49
Tabla V.3 Score de Bureau para cajas municipales de ahorro y créditos	50
Tabla V.4 Impacto financiero en la implementación de un modelo Credit Scoring en una entidad de microfinanzas.....	51
Tabla V.5 A Measurement of the Small Business Credit Gap And The Use of Credit Scoring By Small Financial	52
Tabla V.6 Small Business Credit Scoring: Evidence from Japan.....	53
Tabla V.7 Credit Scoring and the Availability of Small Business Credit in Low- and Moderate-Income Areas	54
Tabla V.8 Modelos de Credit Scoring – Matías Alfredo Gutiérrez Girault.....	55
Tabla V.9 Credit Scoring en instituciones microfinancieras para el caso peruano propuesto por la SBS.....	56
Tabla V.10 Variables Modelo de Credit Scoring Pymes para la medición del riesgo de morosidad de pequeña y mediana empresa	57
Tabla V.11 Modelling Credit risk for SMEs: Evidence from the US Market	58
Tabla V.12 Construcción de Variables	62
Tabla VI.1 Variables Credit Scoring Revolvente	66
Tabla VI.2 Pruebas de coherencia	67
Tabla VI.3 Muestras de operaciones revolventes	70
Tabla VI.4 Proporción de operaciones morosas por muestra	71
Tabla VI.5 Modelo Revolvente.....	81
Tabla VI.6 Principales indicadores Gini - ROC	82

Tabla VI.7 Tabla de clasificación muestra test Pyme Revolvente.....	82
Tabla VI.8 Validación del Modelo en personas naturales y jurídicas	84
Tabla VI.9 Credit Scoring con solo variables del titular de la Pyme.....	86
Tabla VI.10 Correlación de Pearson Credit Scoring Revolvente vs Modelo con información del titular	88
Tabla VI.11 Indicadores de Bondad de Ajuste	89
Tabla VI.12 Intervalos de confianza Credit Scoring revolvente.....	90

CURRICULUM VITAE

ANDY FLORES BARDALES

Magíster en Finanzas, Universidad ESAN. Bachiller en Ingeniería Estadística por la Universidad Nacional de Ingeniería. Especialización en Riesgos, Finanzas Corporativas, Banca de Inversión y Business Intelligence en Universidad ESAN. Diploma en Gestión de Riesgos (IFB - ASBANC). Diploma en Finanzas – Escuela de Empresas de Lima.

Experiencia en Riesgos, Business Intelligence, cobranzas, seguros, gestión de clientes aplicando machine learning, data mining para segmentar portafolios, generar modelos predictivos, Scoring, rating, clv (customer lifetime value), modelos de fuga, modelos de giros con el objetivo de tomar decisiones en base a la data, con inteligencia y entendiendo el comportamiento de los clientes y el mercado.

Formación analítica que permite realizar análisis estratégico, liderazgo de equipos, gestión de proyectos y habilidades personales para lograr los objetivos.

Experiencia Laboral

Especialista Senior - Gestión de clientes y analytics - DinersClub - Perú (diciembre 2018- Actualidad)

- ✓ Desarrollo e implementación de los modelos de seguros y giros de consumo.
- ✓ Campañas dinero en efectivo, incremento de línea para todos los productos y canales.

Especialista Data mining AUNA – Oncosalud (junio 2017- octubre 2018)

- ✓ Desarrollo de iniciativa de ecuación de valor de la compañía: CLV (Customer Lifetime Value) y PLV (Prospect Lifetime Value).
- ✓ Desarrollo de Modelos de Bajas por Morosidad y APC, aplicado a Churn de la compañía.
- ✓ Desarrollo de Segmentación Interna Oncosalud (Platino, Oro y Plata), estrategia para el modelo de atención al cliente de la compañía.

Analista Senior Estrategia de Cobranzas- División de Riesgos - MiBanco (junio 2015- junio 2017)

- ✓ Responsable del Despliegue de la Estrategia de Cobranzas de Credicorp en Mi-banco.
- ✓ Desarrollo e Implementación de Metodología Champion Challenger – Pilotos. Valorización de Cartera.
- ✓ Análisis exploratorios y estadísticos de los diferentes canales de cobranzas, proceso de asignación, proceso de metas, proceso judicial, canales externos, etc.

Sub Gerente Control de Morosidad y Cobranzas - Financiera CrediScotia - Gerencia de Riesgos (julio 2013- noviembre 2014)

- ✓ Desarrollo e implementación de la Estrategia de cobranzas de los segmentos Microcréditos, Consumo, Tarjetas de Crédito y Canal Bodegas.
- ✓ Construcción e implementación de la información analítica para la Cobranza.
- ✓ Calibración de modelos y estrategias.
- ✓ Metodología Champion Challenger – Pilotos.
- ✓ Construcción e implementación del Dashboard de Cobranza.
- ✓ Responsable del Score de Cobranzas y cambios en los Procesos de Castigo, judicial, reprogramaciones, condonación, exoneraciones y políticas de cobranzas.
- ✓ Valorización de la cartera castiga.

Sub Gerente Adj. Modelamiento y Metodologías Gestión de Riesgos - Banco de Crédito BCP - División de Riesgos (noviembre 2010- julio 2013)

- ✓ Desarrollo del Módulo de Comportamiento y Alertas de Empresas en Marcha para el BCP3.
- ✓ Elaboración de la metodología para el Seguimiento de la Familia de Modelos BCP2 (Análisis del Portafolio de Banca Mayorista y Banca Empresa, Grandes, Pequeños, Agrícolas, Constructores y Banca Negocios)

- ✓ Desarrollo metodológico y herramientas para el control de calidad de los modelos de Banca Mayorista y Minorista del BCP. Diagnóstico de Modelos, propuesta para calibración y mejoras estructurales de los modelos.
- ✓ Seguimiento de modelos de Scoring Cualitativos y Cuantitativos.
- ✓ Validación e implementación de modelos de aprobación y pre aprobación de consumo y Pyme (Behavior y Applicant). Comparación de Scores.
- ✓ Control de calidad de modelos para determinar la capacidad predictiva y el seguimiento de la degradación de los modelos de Banca Minorista y Mayorista, propuestas de mejoras.

Analista Data Mining - Banco de Crédito BCP - Business Intelligence (noviembre 2008 - noviembre 2010)

- ✓ Desarrollo de modelos predictivos, segmentación de portafolio y soluciones analíticas para las Gerencias del BCP.
- ✓ Gestión, control y seguimiento de los proyectos analíticos del BCP.
- ✓ Modelo Predictivo para la generación de alertas para lavado de activos / Empresas Fachada
- ✓ Segmentación comercial del Portafolio de Prima AFP
- ✓ Calidad del servicio, modelos de fuga de clientes.

Analista de Proyectos - Fundación Desarrollo Integral de Nuevo Pachacútec (julio 2004 – noviembre 2008)

Ejecución del Plan Director del Proyecto Universidad Católica de Pachacútec, implementación de las carreras de electricidad, administración de empresas y cocina. Responsable de la implementación y evaluación de proyectos, responsable de las finanzas, aplique análisis estadísticos multivariados y modelos estadísticos.

RESUMEN EJECUTIVO

El objetivo de esta tesis es desarrollar un modelo de Credit Scoring para el segmento Pyme, por lo cual se ha revisado las principales metodologías: Paramétrica, No paramétrica y Semi- Paramétrica aplicadas al Credit Scoring. Se aprecia que el modelo logístico (paramétrico) es usado por las instituciones financieras por su fácil interpretación, bondad de ajuste matemático y sobre todo porque establece una línea de comparación con las diferentes metodologías. No hay una metodología óptima, mucho depende de la calidad de la información y de las variables que ingresen al modelo.

El modelo de Credit Scoring se desarrolla para el producto capital de trabajo en la modalidad revolviente por ser el que presenta mayor demanda. Contiene información del titular y socios o intervinientes. Se determinan las características o atributos más importantes para discriminar el riesgo por lo cual las variables a considerar son: (i) el saldo activo de los últimos tres meses entre el saldo activo de los últimos 24 meses, (ii) el número de empresas que reportan al titular, (iii) los sobregiros, (iv) la evolución de la deuda últimos tres reportes entre los últimos 24 meses, (v) el pasivo entre el patrimonio, (vi) el capital de trabajo, (vii) el patrimonio, (viii) el atraso máximo en los últimos 12 meses, (ix) el máximo saldo vencido en los últimos 12 meses, (x) la diferencia de meses con buen comportamiento y mal comportamiento, (xi) la máxima clasificación en el sistema financiero los últimos 24 meses, (xii) los pasivos, (xiii) los meses de constitución de la empresa, (xiv) la máxima antigüedad en pasivos o activos y (xv) la zona geográfica.

El modelo propuesto a diferencia de las aplicaciones revisadas, es un modelo personalizado al riesgo de la institución financiera, de admisión de clientes, muy preciso en la aprobación de créditos y se aplica sólo a personas que cuentan con experiencia previa en el sistema financiero. Se considera la información del titular de la Pyme y socios. Las aplicaciones revisadas no diferencian entre clientes que usan las modalidades de crédito revolviente y no revolviente, son modelos genéricos. El modelo se construye con información interna de la institución financiera, así como información del Reporte Crediticio Consolidado (RCC), por lo tanto, no se considera como un modelo

genérico tipo buró, en su construcción no se utilizan otros modelos como input, lo que si sucede con algunos modelos revisados en las aplicaciones para Pyme.

Las variables del Credit Scoring propuesto presentan una profundidad histórica de 24 meses, sin embargo, las aplicaciones revisadas no tienen una profundidad histórica tan amplia, la ventana de predicción del Credit Scoring propuesto es de 12 meses lo que garantiza estimaciones estables. En los modelos revisados se observa que no se explota la información del titular y los socios de la Pyme, en el Credit Scoring propuesto se considera dicha información y además se diferencia a las personas naturales y jurídicas. Actualmente existen muchas aplicaciones de Credit Scoring, sin embargo, omiten realizar pruebas de estabilidad predictiva a través del uso de intervalos de confianza en sus indicadores Gini y ROC, lo que implica que se puede estar sobreestimando o subestimando el riesgo de crédito. El Credit Scoring propuesto garantiza su performance mediante la simulación de Monte Carlo para determinar los intervalos de confianza para los indicadores en mención. Otro aspecto relevante en el Credit Scoring propuesto es que no se ha detectado concentración en el peso de las variables utilizadas, dichas variables son independientes entre sí, no se aprecia multicolinealidad y discriminan el riesgo.

La metodología utilizada es la regresión logística debido a que se considera una técnica referente para discriminar el riesgo de crédito, es intuitiva, fácil de entender e implementar. En la muestra de entrenamiento (intramuestral) del modelo se obtiene un indicador Gini de 51.5% y en la muestra test (fuera de muestra) se obtiene un indicador Gini de 48%.

Un aporte relevante del modelo es la estabilidad probada intramuestral y fuera de muestra, ya que en este sentido se ha generado 4,000 simulaciones para el cálculo de los indicadores Gini y ROC determinando que sus valores se encuentran dentro de los intervalos de confianza (al 95% y al 99%). Por lo tanto, el modelo propuesto garantiza un alto rendimiento predictivo ya que ha sido sometido a pruebas de estabilidad predictiva (Simulación de Monte Carlo- Bootstrap). En las aplicaciones revisadas se pudo apreciar que no se establece intervalos de confianza para la estimación de los indicadores de predicción intramuestral y fuera de muestra.

La tasa correcta de clasificación del modelo desarrollado es de 79.7% en la muestra test (fuera de muestra) lo que confirma la estabilidad del modelo.

Se aprecia en el modelo Credit Scoring que existen diferencias significativas entre persona natural y jurídica. El modelo en mención se aplica solo a personas naturales y se observa que las variables que toman mayor peso o importancia son las variables de la experiencia en el sistema financiero (empresas donde reportan al titular, evolución de deudas y clasificación), saldos pasivos, meses de constitución de la Pyme y la zona geográfica. Cuando se aplica el modelo a solo personas jurídicas el peso de las variables se inclina al sobregiro, patrimonio, saldos pasivos y meses de constitución de la Pyme. Las diferencias existentes entre los grupos de personas naturales y jurídicas son confirmadas por el modelo agregado, en el que la variable ficticia que hace referencia a los grupos (natural y jurídica) es estadísticamente significativa. Finalmente se considera la variable “persona natural jurídica” en el modelo como una variable adicional y se observa que es estadísticamente significativa con un p-value menor a 0.05 e incrementa el indicador Gini a 51,6%.

Uno de los aportes relevantes de la tesis es que permite evidenciar que el riesgo de la Pyme es el riesgo del titular de la Pyme. Se genera un modelo con sólo variables del titular y se obtiene un indicador Gini de 49.9%, el cual se compara con el indicador Gini de 51.5% obtenido del modelo Credit Scoring revolvente. Se observa que la diferencia de ambos indicadores es muy pequeña, lo que intuitivamente refleja que el aporte de los socios o intervinientes no es relevante. Al cruzar ambos modelos en un diagrama de dispersión se observa que la correlación es 0.85. Esto quiere decir que ambos modelos tienen una relación directamente proporcional y se aprecia que cuando las probabilidades de incumplimiento no son elevadas prácticamente se puede utilizar cualquiera de los modelos para calcular el riesgo de crédito de la Pyme.

Por lo tanto, ambos modelos Credit Scoring (con sólo variables del titular) y el Credit Scoring revolvente (con variables del titular y socios) presentan alto poder de predicción y están correlacionados, por lo que se muestra evidencia que el riesgo de la Pyme es el riesgo del titular. Sin embargo, se sugiere realizar un análisis con mayor profundidad.

CAPÍTULO I. INTRODUCCIÓN

1.1. Introducción

Según la Asociación de Bancos del Perú - Asbanc (2017) las pequeñas y medianas empresas (Pymes) son unidades productivas con una importante presencia en América Latina, representan un sector importante para la economía y la banca. Según el reciente informe de la Federación Latinoamericana de Bancos - Felaban (2016) el 87% de los bancos de la región cuentan con una sección especializada en atender los servicios y requerimientos de las Pymes.

Para Asbanc (2017) las Pymes son muy importantes para los bancos de la región, pues se estima que el 96% de los bancos consideran que atender a este segmento es parte de su estrategia integral de negocio. Asimismo, la mayor parte de ellos consideran que las Pymes representan una oportunidad de negocio. Sin embargo, hay obstáculos que se presentan durante el proceso de otorgamiento del crédito para las Pymes. Los bancos consideran que todavía existen excesivos requerimientos de información por parte de las áreas de negocio y la dificultad en la recolección de información y procesamiento de aplicaciones Crediticias.

Según los balances de comprobación del BCR (2018) la Pyme tiene una tasa de incumplimiento que oscila entre 9.5% y 9.9%, en los últimos 3 años, esto revela el mayor riesgo de este segmento.

En la tabla I.1, se observa el ratio de morosidad del sistema financiero, la banca y no banca (conformada por financieras, cajas municipales y cajas rurales) donde el segmento Pyme presenta ratio de morosidad que oscila entre 9.5% y 9.9% en los últimos 3 años, en el sistema financiero (incluye banca y no banca).

Tabla I.1 Ratio de morosidad del sistema financiero

Ratio de morosidad del sistema financiero									
	Sistema financiero			Banca			No Banca		
	Mar-16	Mar-17	Mar-18	Mar-16	Mar-17	Mar-18	Mar-16	Mar-17	Mar-18
Total	4.2	4.6	4.9	3.8	4.3	4.5	7.9	7.3	8
A Empresas:	4.3	4.7	5.1	3.7	4.2	4.5	9.4	8.4	9.1
Corporativos	0	0.1	0.2	0	0.1	0.2	0	0	0
Grandes	1.6	2	2.5	1.6	2	2.5	7.3	8.1	5.9
Medianas	8.2	9.8	10.8	8	9.8	10.8	10.9	10	11.6
Pyme	9.9	9.5	9.6	10.3	10.6	10.3	9.5	8.3	8.9
A Hogares	4	4.5	4.7	3.9	4.4	4.5	5	5.4	6.2
Consumo	5.2	5.5	5.4	5.1	5.4	5.2	5.4	5.7	6.5
Hipotecarios	2.6	3.2	3.7	2.6	3.2	3.7	2.7	3.7	4.1

Fuente: BCR, Balances de comprobación (2018) y Elaboración propia

En el Perú, desde hace algunos años la banca Retail viene incursionando rápidamente en las microfinanzas. Este proceso es llamado downscaling y consiste en que los grandes bancos (Bcp, Bbva, Scotiabank, Interbank, entre otros) proveen préstamos o créditos a microempresarios. Esta estrategia de negocio se lleva a cabo mediante la compra de instituciones microfinancieras que están en el mercado o mediante la creación de un área especializada dentro del banco que desarrolle y ponga en producción la tecnología Crediticia para conceder créditos a la Pyme. La gran mayoría de estos créditos financian el capital de trabajo y en segundo lugar los activos fijos (Morán, 2010).

Las soluciones financieras que se ofrecen a las Pymes son microcréditos, capital de trabajo en la modalidad revolviente y no revolviente, activos fijos inmuebles y mueble, carta fianza, certificados bancarios, créditos multipropósito, descuento de facturas negociable, descuento de letras, factoring electrónico, leasing y letras de cobranza en garantía. Dichos productos se ofrecen a personas naturales con negocio propio o persona jurídica (EIRL, SAC, SRL, SA).

Las instituciones financieras (bancos, financieras, cajas, edPymes y otros) vienen generando modelos estadísticos, los cuales son desarrollados muchas veces por consultoras internacionales y también por los analistas de riesgos, estos modelos permiten gestionar el riesgo de crédito e implementar diversas estrategias en los diferentes productos, como préstamos personales, vehiculares, tarjetas de crédito, hipotecarios y también en la Pyme con el objetivo de seleccionar mejor a sus clientes y disminuir su proporción de cartera atrasada.

En el Perú, las entidades financieras vienen adoptando modelos internos para gestionar el riesgo de crédito en el marco del acuerdo de Basilea, en la bibliografía se observa aplicaciones de Credit Scoring en diferentes realidades donde solo cuentan con información de centrales de riesgo e información interna de la institución financiera. Es de público conocimiento que la Superintendencia de Banca y Seguros SBS desde el año 2001, consolida la información de deudas del sistema financiero mediante el Reporte Crediticio Consolidado (RCC) para luego redistribuir a todas las entidades financieras. Con la información de la RCC y con información interna de la institución financiera se desarrolla un modelo Credit Scoring para otorgar préstamos a clientes del segmento Pyme.

En el Perú, la Pyme presenta un elevado riesgo de crédito por el alto grado de informalidad del negocio, la estructura de su composición gerencial, por la naturaleza del negocio que es familiar y es dirigido por personas con formación educativa básica. Sumado a que muchas veces operan con las mínimas condiciones de trabajo y con flujos de caja muy volátiles. Además, muchas veces por ahorrar costos prefieren seguir en la informalidad.

Se desarrolla un Credit Scoring, asumiendo la realidad del negocio y definiendo las características más predictivas que nos ayuden a identificar el riesgo en las operaciones de crédito. El presente trabajo presenta una propuesta simple, pero sólida en sus fundamentos estadísticos al construir un modelo de Credit Scoring para Pymes, utilizando una regresión logística para calcular la probabilidad de que un cliente pague o no el crédito.

Uno de los aportes relevantes de la tesis es que permite evidenciar que el riesgo de la Pyme es el riesgo del titular de la Pyme. Se genera un modelo con sólo variables del titular y se obtiene un indicador Gini de 49.9%, el cual se compara con el indicador Gini de 51.5% obtenido del modelo Credit Scoring revolving. Se observa que la diferencia de ambos indicadores es muy pequeña, lo que intuitivamente refleja que el aporte de los socios o intervinientes no es relevante. Al cruzar ambos modelos en un diagrama de dispersión se observa que la correlación es 0.85. Esto quiere decir que ambos modelos tienen una relación directamente proporcional y se aprecia que cuando las probabilidades de incumplimiento no son elevadas prácticamente se puede utilizar cualquiera de los modelos para calcular el riesgo de crédito de la Pyme.

La propuesta brinda una perspectiva diferente frente a los modelos de admisión para Pyme ya que pone énfasis en el riesgo del titular de la Pyme.

1.2. Objetivo

Desarrollar un modelo Credit Scoring para el otorgamiento de préstamos a clientes del segmento Pyme de una institución financiera, probando la estabilidad del modelo dentro y fuera de muestra, validándolo en personas naturales y jurídicas y mostrando que el riesgo de la Pyme está fuertemente correlacionado con el riesgo del titular.

1.3. Objetivos específicos

- ✓ Revisión crítica de los modelos o metodologías de Credit Scoring.
- ✓ Desarrollar un modelo Credit Scoring para el producto Pyme “capital de trabajo revolvente”, por ser el que tiene mayor demanda.
- ✓ Evaluar las principales características o variables que influyen en el incumplimiento de pago en el producto Pyme “capital de trabajo”, usando un modelo de Credit Scoring.
- ✓ Evaluar si los resultados del modelo Credit Scoring presenta diferencias significativas entre personas naturales y jurídicas.
- ✓ Evaluar si el riesgo de la Pyme está correlacionado con el riesgo del titular.

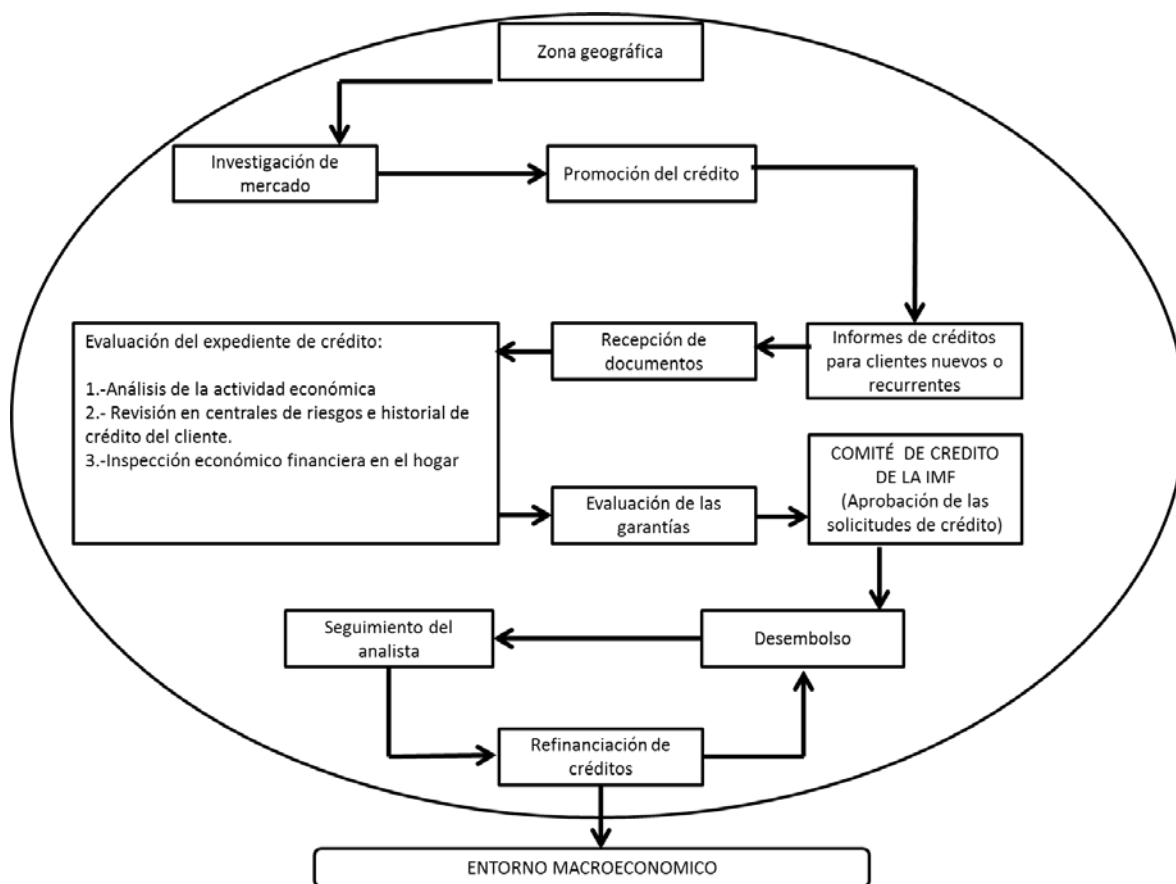
CAPÍTULO II. CREDIT SCORING

En este capítulo se muestra el proceso Crediticio, se define que es un Credit Scoring, cuál es el propósito de un Credit Scoring, como se aplica un Credit Scoring, ventajas y desventajas, tipos de Credit Scoring y cuál es su relación con el nuevo acuerdo de capital Basilea.

2.1. Proceso Crediticio

Según Rayo et al. (2010) desarrollan el proceso de promoción, evaluación y concesión de un microcrédito que se resume en la figura II.1:

Figura II.1 Proceso de promoción, evaluación y concesión de un microcrédito



Fuente: Rayo et al., Un modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II (2010) y Elaboración propia

En la Figura II.1 se presentan las fases del proceso Crediticio desde que el crédito es promovido hasta su aprobación y entrega. Se debe determinar en cada una de las etapas del proceso las variables explicativas (cliente, operación y macroeconómicas) como también el objetivo a pronosticar. Según Rayo et al. (2010), el proceso para la concesión de un crédito tiene las siguientes etapas:

a) Investigación de mercado y promoción del crédito

Se identifica la agencia donde se realizará el estudio de mercado y la promoción del crédito. La entidad financiera se dedicará a informar a las personas naturales y jurídicas sobre los créditos que brinda, con el objetivo de ofrecer créditos para satisfacción de las necesidades del cliente. En esta etapa del proceso identificamos la variable zona geográfica, según Rayo et al. (2010).

b) Informes de créditos para clientes nuevos y recurrentes

La institución, informa y orienta al cliente sobre los productos y servicios que ofrece. La empresa brinda asesoría financiera para el cliente y, también ayuda a los clientes mediante asesoría financiera. Todo esto con la finalidad de presentar al cliente el producto financiero más adecuado para su necesidad. Las variables que aportarían al Credit Scoring en esta etapa serían tiempo del prestatario como cliente de la entidad, créditos entregados con anterioridad y créditos rechazados con anterioridad, según Rayo et al. (2010).

c) Evaluación del expediente del crédito

Esta actividad la realiza el analista de crédito, quien evalúa los expedientes crediticios para presentarlos al comité de créditos, quienes aprueban o rechazan el pedido. Las fases a realizar son tres:

Evaluación de la actividad económica desarrollada por el cliente, las variables para el Credit Scoring serían: sector y destino del microcrédito.

Se revisa el comportamiento de pago en el RCC, en las centrales de riesgo, en las bases de datos internas de la institución financiera. De esta etapa se sacan las variables de morosidad y los indicadores de cumplimiento de pago.

Finalmente se hace una visita al negocio o microempresa donde se hace una inspección económico financiera y las variables que se recogen son de tipo patrimonial, es decir se obtienen todos los ratios económicos y financieros y la información personal de los solicitantes como el sexo, la edad, el estado civil, la situación laboral entre otras Rayo et al. (2010).

d) Evaluación de las garantías

Antes de entregar el expediente al comité de créditos, el analista y el jefe de asesoría legal, evalúan las garantías que el cliente ofrece. En esta etapa se incluye la variable garantía Rayo et al. (2010). Sin embargo, hay que tener en cuenta que las garantías deben ser reales, es decir deposito en cuenta, inmuebles, etc. No se puede aceptar como garantía las herramientas de trabajo ya que no garantizan su rápida venta y podrían terminar generando costos para la institución financiera. En el modelo propuesto la variable garantías reales no es relevante, ya que no se cuenta con dicha información y además los clientes que tienen garantías reales solo cubren el 1% de la base de datos y el 99% son vacíos o no tienen garantías reales.

e) Aprobación de la solicitud del crédito

En esta etapa del proceso se aprueba o rechaza el crédito. Las variables que se toman en cuenta son tipo de operación, moneda, duración, tasa de interés mensual, cantidad rechazada por la institución, pronóstico del analista sobre la situación del crédito a su vencimiento Rayo et al. (2010).

f) Variables macroeconómicas

El ciclo económico impacta en el desarrollo del modelo, se debe tener en cuenta las variables macroeconómicas producto bruto interno, índice de precios al consumidor, índice de empleo, tasa de cambio, tipos de interés, índice general bursátil, tarifa de agua, tarifa de luz y tarifa de teléfono entre otras. Dichas variables deben estar expresadas como tasas de variación anualizada y durante el periodo de vigencia del crédito.

Las variables independientes que se muestran en el proceso Crediticio descrito ayudaran a predecir la variable dependiente, en el caso del proceso Crediticio mostrado la variable a predecir sería el incumplimiento de pago el cliente Rayo et al. (2010).

2.2. Definición de Credit Scoring

Según Thomas, Edelman y N. Crook (2004), se define Credit Scoring como el conjunto de modelos de decisión y técnicas que ayudan a las entidades financieras en la entrega de créditos de consumo. A partir de estas técnicas se decide quién recibirá el crédito, cuanto crédito debería obtener y qué estrategias operativas mejorarán la rentabilidad de las entidades financieras sobre los créditos aprobados.

Los modelos de Credit Scoring estiman el riesgo de los préstamos para un consumidor en particular.

La entidad financiera tomará dos tipos de decisiones, siendo el primero el de otorgar crédito a un nuevo solicitante y, segundo, el cómo tratar con las solicitudes que se vuelven clientes del portafolio, incluso si aumentar o no sus límites de crédito. Las técnicas que ayudan a la primera decisión se llaman Credit Scoring y las técnicas que ayudan a la segunda decisión se llaman Credit Scoring de comportamiento o behavior y se utilizan cuando un crédito es parte de la cartera de la institución financiera.

Lawrence y Solomon (2002) definen al Credit Scoring como el uso de técnicas estadísticas para identificar, ordenar y clasificar a un atractivo o potencial cliente o a los clientes actuales de una empresa.

Hand & Henley (1997), citado por Gutiérrez (2007) sostiene que: el Credit Scoring se puede definir como: “métodos estadísticos utilizados para clasificar a los solicitantes de crédito, o incluso a quienes ya son clientes de la entidad evaluadora, entre las clases de riesgo bueno y malo”.

Adicionalmente, Gutiérrez (2007) comenta: “A pesar de la proliferación de los modelos de Credit Scoring, el juicio humano (o juicio del analista) continúa siendo utilizado en la admisión de los créditos, en algunos casos expresado como un conjunto de reglas que la entidad aplica de manera sistemática para filtrar solicitudes o deudores”.

Las reglas de negocio de la entidad también conocidas como política de créditos, evalúan si es posible aprobar o denegar un crédito para los clientes nuevos o no bancarizados sin experiencia en el sistema financiero, mediante estas reglas se mide nivel de ingresos del cliente, si tiene casa propia, automóvil, carga familiar, si es miembro de algún club, etc. En este sentido Gutiérrez (2007) afirma que ambas metodologías Credit Scoring y juicio experto pueden convivir y ser complementarias. Para los clientes nuevos, pasado un tiempo se puede aplicar el Credit Scoring en la etapa de seguimiento del crédito, porque ya se conoce el comportamiento de pagos del cliente. En el presente trabajo la institución financiera donde se aplica el Credit Scoring no concede créditos a clientes sin experiencia crediticia, ya que su política de créditos no le permite.

Según la experiencia del autor de la presente tesis, se denomina Credit Scoring a un conjunto de técnicas multivariadas (regresión logística, probit, tobit, redes neuronales, análisis discriminante, árboles de decisión, programación lineal, entre otras) así como a las técnicas machine learning, Deep learning e inteligencia artificial entre otras, que explican mediante un conjunto de variables o características, a un evento que puede tomar dos valores 1 si es default (si el cliente no paga el crédito en un determinado periodo de tiempo) y 0 si el cliente cumple con el pago del crédito.

Las variables independientes deben de tener por lo menos un año de profundidad histórica en la data, mientras que la variable dependiente se debe observar en una ventana de tiempo de 1, 3, 6 y 12 meses de acuerdo a la necesidad del negocio, por ejemplo, para modelos de cobranza se necesita que la variable dependiente tenga solo un mes de ventana de observación, ya que se necesita estimar que clientes caen en mora en los siguientes 30 días. Cuando se desarrolla un modelo se busca garantizar su estabilidad, es por ello que la definición de la variable dependiente debe ser muy precisa ya que todos los análisis se centran en explicar dicha variable e incluso las muestras de entrenamiento y test dependen de cierta forma de esta variable.

La ventana de predicción del modelo dependerá del objetivo a modelar, por ejemplo, en el caso de un modelo de admisión de nuevos clientes se trabaja con fotos mensuales, cohortes o cosechas cuyo objetivo a pronosticar, target o default es calcular la probabilidad que un cliente caiga en atraso mayor o igual a 60 días en los siguientes 12 meses, caso contrario no es default, si el cliente no se atrasa más de 60 días en los siguientes 12 meses.

2.3. Propósito del Credit Scoring

El propósito del Credit Scoring, es predecir si un cliente al otorgarle un préstamo, pagará el préstamo de manera oportuna o no (Banasik et al., 2003). Las instituciones financieras utilizan el Credit Scoring cuando evalúan las solicitudes de préstamos (Mester, 1997) mediante la evaluación del riesgo de los préstamos a diferentes consumidores (Bellotti y Crook, 2009). La predicción y la evaluación son necesarias para determinar cuál de los solicitantes, tiene la capacidad de pagar el préstamo y por lo tanto, se le puede otorgar un préstamo (Banasik et al., 2003).

Los modelos de Credit Scoring no solo ayudan en la aprobación de préstamos, sino también en la determinación de los precios y el control de los préstamos. Los prestatarios potenciales se clasifican de acuerdo con su probabilidad de impago sobre un préstamo relevante de acuerdo con los datos utilizados en la solicitud de crédito y la referencia Crediticia del individuo o la empresa (Bellotti y Crook, 2009). El monto del crédito, la administración del crédito y los riesgos de la cartera de créditos se pueden identificar con el uso del Credit Scoring (Turvey & Brown, 1990).

Los modelos de Credit Scoring tienen el potencial de reducir la variabilidad en las decisiones de crédito y agregar eficiencias en las evaluaciones de riesgo Crediticio (Limsombunchai, Gan & Lee, 2005).

2.4. Aplicaciones del Credit Scoring

El Credit Scoring se aplica en la aprobación de préstamos de consumo, tarjetas de crédito, hipotecas, vehicular, Pyme, corporativos bajo su denominación de Rating, etc.

Para aprobar o rechazar una solicitud de crédito se fija un punto de corte, en la presente tesis es un nivel de probabilidad, las operaciones que superen el punto de corte se aprueban caso contrario se rechazan. En tal punto de corte se toma en cuenta el grado de riesgo a asumir, es decir, la intensidad de riesgo que el banco está dispuesto a aceptar y los beneficios que la institución quiere obtener.

En la banca el Credit Scoring se aplica para:

- ✓ Aprobar o rechazar una solicitud de crédito.
- ✓ Direccionar las ofertas a los mejores prospectos.
- ✓ Administrar el portafolio de créditos, mediante los modelos de seguimiento, proactivos o behavior Scoring.
- ✓ Evitar la fuga de clientes de la compañía.
- ✓ Segmentar y perfilar la cartera de clientes.
- ✓ Realizar incrementos o decrementos de línea de una tarjeta de crédito, disposición de efectivo, préstamos personales y en cuentas corrientes.
- ✓ Analizar la rentabilidad de los clientes y los potenciales clientes.
- ✓ Fijar precios de los productos de acuerdo al nivel de riesgo de los clientes.
- ✓ Seguimiento y la gestión de las cobranzas.
- ✓ En la estrategia de oferta de nuevos productos.

2.5. Ventajas y desventajas del Credit Scoring

Para describir las ventajas y desventajas se toma lo enunciado por Schreiner (2002).

1. Cuantifica el riesgo como una probabilidad y es consistente y explícito para un periodo de tiempo determinado.
2. Proceso más eficiente en el tiempo traduciéndose en menos costos.
3. Cada crédito está basado en un contexto probabilístico, lo que permite un control estadístico del portafolio.
4. Permite evaluar a cada uno de los clientes.
5. Permite reducir el tiempo invertido en cobranza por parte de los analistas de crédito.
6. Son procesos que pueden tomar bastante tiempo, porque se tiene que construir y probar muchas variables para identificar a las más predictivas.
7. Son modelos de predictibilidad limitada, porque el comportamiento de los clientes es cambiante, también hay una dependencia de los ciclos económicos y la transformación digital.
8. La base de datos casi siempre contiene información no precisa o aleatoria por lo tanto mientras esas perturbaciones no sean demasiado fuertes, el modelo puede captar las señales de riesgo que emiten las características presentes en dichas bases de datos.
9. Se hace necesario contar con un área especializada para su monitoreo y desarrollo.
10. Para el caso de un nuevo cliente que no tenga información se establecen políticas de excepción para conceder un crédito ya que algunas veces no se puede aplicar el Credit Scoring. Luego de un tiempo en la medida que el cliente muestra un comportamiento saludable en sus pagos se le puede mejorar las condiciones iniciales del crédito.

2.6. Tipos de Credit Scoring

Los tipos de Credit Scoring son los siguientes: Lawrence y Salomon (2002)

- ✓ **Applicant o de admisión**, son los modelos que se utilizan para la admisión de un crédito, es decir en el momento de la solicitud de una nueva operación. También sirven para pre aprobación de clientes.
- ✓ **Behavior**, o modelos de comportamiento estos modelos se utilizan en la gestión de portafolio para el seguimiento de los créditos al día y con retraso. También se utilizan en las campañas de incremento o decremento de línea para las tarjetas de crédito, disposición de efectivo a tasas especiales, préstamos personales, compras de deuda, líneas paralelas, créditos pre aprobados, etc.
- ✓ **Buró**, estos modelos se construyen con la información del sistema financiero, información que es compartida por todos los bancos peruanos y las centrales de riesgo. Se usan para clasificar a un cliente ya sean nuevos o clientes actuales son robustos ya que utilizan información como mínimo de 5 años de historia.
- ✓ **Collection**: Estos modelos se utilizan en la gestión de cobranzas generalmente se desarrollan en los diferentes tramos de atraso como por ejemplo la mora a 30, 60 y 90 días. También existen los modelos de recuperación para los créditos castigados.
- ✓ **Rating**: Son modelos Credit Scoring aplicados al segmento de banca negocios, banca empresa y corporativos. Casi siempre utilizan información financiera de los clientes, la tecnología Crediticia, el uso de los sobregiros, garantías, cartas fianzas, cheques, entre otros instrumentos financieros es bastante frecuente.

2.7. Acuerdos de Capital Basilea I, II, III y el Credit Scoring

El nuevo acuerdo de capital Basilea I, surge como una necesidad de establecer estándares en la banca a nivel internacional, su inicio se remonta al año 1988 y entra en vigencia en 1992, el objetivo del acuerdo de capital es salvaguardar a los depositantes y exige a los bancos un nivel de capital mínimo, para mitigar el riesgo de crédito, mercado y tipo de cambio. El mínimo capital solicitado es de 8% Bis (2004).

Luego de unos años surge Basilea II y propone tres pilares, el pilar 1 exige un nivel de capital mínimo, todos los reguladores internacionales exigen a las instituciones financieras desarrollar modelos estadísticos para la gestión del riesgo de crédito. Basilea II brinda el marco metodológico para estimar la probabilidad de default, la pérdida dado el incumplimiento y la exposición dado el incumplimiento. El pilar 2 propone la supervisión de la gestión del capital de los socios, se puede solicitar mayor capital a los accionistas, con el fin de cubrir los riesgos de inherentes al negocio. El Pilar 3 trata sobre la disciplina de mercado, es necesario que las entidades financieras sean transparentes con la información cualitativa y cuantitativa SBS (2008) & Bis (2004).

La crisis financiera internacional del año 2008, obligo a los bancos y entidades financieras a dar importancia a la administración de la liquidez de los diferentes agentes financieros. Debido a ello surge Basilea III, con un enfoque de Riesgo de Liquidez Bis (2017).

El comité de Basilea recomienda aplicar modelos estadísticos y matemáticos para generar modelos metodologías de Credit Scoring, rating, probabilidad de default, pérdida dado el incumplimiento, exposición dado el incumplimiento, comparaciones mediante backtesting, stresstesting, benchmarking, seguimientos, calibraciones y evaluaciones periódicas a los modelos.

CAPÍTULO III. PYMES

En este capítulo se revisa como se define a la Pyme, cuáles son sus características, los tipos de créditos, categorías de clasificación en el sistema financiero, que productos brinda el sistema financiero a la Pyme y seleccionaremos a uno de los productos más relevantes para desarrollar un Credit Scoring.

El Congreso de la República (2013) promulgó la Ley N° 30056 “Ley que modifica diversas leyes para facilitar la inversión, impulsar el desarrollo productivo y el crecimiento empresarial”. Se brinda un marco normativo para promocionar la competitividad, ayudar en la formalización e impulsar el avance de las micro, pequeñas y medianas empresas. Se modifican algunas leyes como la “Ley MYPE” (D.S. N° 007-2008-TR.) “Texto Único Ordenado de la Ley de Promoción de la Competitividad, Formalización y Desarrollo de la Micro y Pequeña Empresa y del Acceso al Empleo Decente” Congreso de la República (2008). La ley busca promover e impulsar a las Pymes para aumentar los niveles de empleo, de producción, asegurar rentabilidad, incrementar Producto Bruto Interno, ampliar mercados internos y externos, así como mejorar la recaudación de la Sunat Congreso de la República (2013).

3.1. Definición de Pyme

Según el Congreso de la República (2003), se define a la PYME (Pequeña y Micro Empresa) como la unidad económica ya sea persona natural o jurídica, bajo cualquier forma de organización contemplada en la ley vigente, que tiene como fin desarrollar actividades de extracción, transformación, producción, comercialización de bienes o prestación de servicios.

En la tabla IV.1 se resumen las características que tiene la Pyme:

Tabla IV.1 Definición de Pyme

LEY MYPE D.S. N° 007- 2008 - TR			LEY N° 30056	
	Ventas Anuales	Trabajadores	Ventas Anuales	Trabajadores
Microempresa	Hasta 150 UIT	1 a 10	Hasta 150 UIT	No hay límites
Pequeña Em- presa	Hasta 1,700 UIT	1 a 100	Entre 150 UIT y 1,700 UIT	No hay límites
Mediana Em- presa			Más de 1,700 UIT y hasta 2,300 UIT	No hay límites

Fuente: Congreso de la República. Ley 28015 Ley de Promoción y Formalización de la Micro y Pequeña Empresa (2013) y Elaboración propia

3.2. Característica de la Pyme

La Pyme tiene las siguientes características:

A. Número de trabajadores:

- La pequeña empresa no tiene límite de trabajadores (anteriormente era de 1 a 100 trabajadores)
- La mediana empresa no tiene límite de trabajadores.

B. Ventas anuales:

- La pequeña empresa vende más de 150 Unidades Impositivas Tributarias – UIT y hasta 1700 UIT
- La mediana empresa vende más de 1,700 UIT y hasta 2,300 UIT.

3.3. Tipo de Créditos (Pyme)

A continuación, se presenta algunas definiciones del tipo de crédito definido por la SBS (2009) mediante la Resolución SBS N° 808 – 2003 y la Resolución N° 14353:

3.3.1 Créditos a Pequeñas Empresas

Son los créditos destinados a producir, comercializar o brindar servicios, concedidos a personas naturales o jurídicas, cuya deuda total en el sistema oscila entre S/ 20,000 y S/ 300,000 (sin incluir deudas hipotecarias destinadas a vivienda) en los últimos seis (6) meses.

Si después, la deuda total en el sistema supera los S/ 300,000 (sin incluir deudas hipotecarias destinadas a vivienda) por seis (6) meses consecutivos, los créditos serán reclasificados como créditos a medianas empresas. Pero si, la deuda total en el sistema es inferior a S/ 20,000 (sin incluir deudas hipotecarias destinadas a vivienda) por seis (6) meses consecutivos, los créditos serán reclasificados como créditos a microempresas SBS (2009).

3.3.2 Créditos a Microempresas

Son los créditos destinados a producir, comercializar o brindar servicios, concedidos a personas naturales o jurídicas, cuya deuda total en el sistema es menor o igual a S/ 20,000 (sin incluir deudas hipotecarias destinadas a vivienda) en los últimos seis (6) meses.

Si después, la deuda total en el sistema supera los S/ 20,000 (sin incluir deudas hipotecarias destinadas a vivienda) por seis (6) meses consecutivos, los créditos serán reclasificados al tipo de crédito que corresponda, según su deuda SBS (2009).

3.4. Categorías de Clasificación Crediticia del Deudor de Créditos a Pequeñas Empresas, a Microempresas, de Consumo Revolvente y No Revolvente

Las categorías definidas por la SBS (2018) son: (i) normal si tienen un atraso hasta 8 días calendario, (ii) con problemas potenciales si tienen atrasos entre 9 a 30 días calendario, (iii) deficiente si tienen atraso entre 31 y 60 días calendario, (iv) dudoso si tienen atrasos entre 61 y 120 días calendario y (v) perdida si tienen atrasos superiores a los 120 días calendarios. Dichas categorías están detalladas en la Resolución N° 11356 de la SBS.

3.5. Productos ofrecidos a las Pymes

Según el Ministerio de Producción (2010) los productos financieros que se ofrecen a personas naturales con negocio propio o persona jurídica (E.I.R.L., S.A.C., S.R.L., S.A) son los siguientes:

Microcréditos: Son prestamos pequeños para ayudar a poner en marcha pequeñas empresas, las instituciones financieras como Mibanco, Crediscotia, EdPymes y las cajas son especializadas en este tipo de ofertas.

Capital de trabajo, producto ofrecido a la micro y pequeña empresa (persona natural o jurídica) generalmente son créditos de considerable saldo y son brindados por todas las instituciones financieras, bancos, financieras, EdPymes y cajas. La modalidad de préstamos es revolvante y no revolvante y se utilizan para incrementar el capital de trabajo y así financiar las mercaderías, materias primas, insumos o materiales.

Activos fijos para Pymes, este producto tiene dos modalidades de préstamos, por un lado, los activos fijos mueble que financian la compra de máquinas, equipos y unidades de transporte. Por otro lado, los activos fijo inmueble que financian la compra de locales comerciales tiendas o talleres.

Leasing, para renovación de activos, ampliar plantas, líneas de producción, entre otros.

Factoring, es un producto que se utiliza para obtener dinero a cambio de facturas por montos pendientes de cobrar. El BCP, BBVA Banco Continental, Interbank, Mibanco y Scotiabank ofertan este producto.

Descuento comercial, esta modalidad consiste en ceder a favor del banco el cobro de una factura, con la condición de que el banco anticipe la deuda menos los intereses y gastos. Este producto brinda liquidez inmediata y su trámite es sencillo, se necesita un pagaré una letra o un recibo, dicho producto es ofrecido por la mayoría de los bancos.

Otros productos que ofrecen las instituciones financieras a las Pymes son renting, prestamos subvencionados, lease-back y rent-back, etc.

3.6. Indicadores Pyme

Se mencionan los siguientes indicadores:

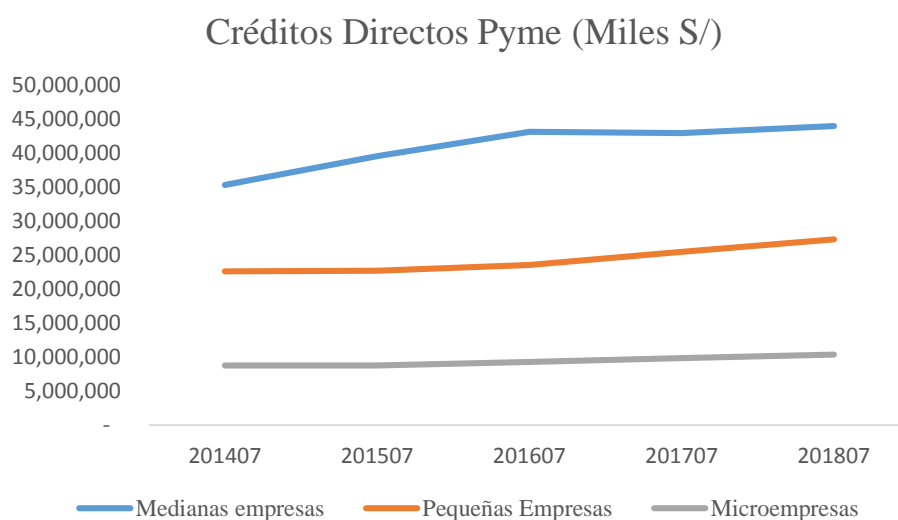
Créditos Directos: Según la SBS (2015) se define: “Representa los financiamientos que, bajo cualquier modalidad, las empresas del sistema financiero otorguen a sus clientes, originando a cargo de éstos la obligación de entregar un monto de dinero determinado. Corresponde a la suma de los créditos vigentes, reestructurados, refinanciados, vencidos y en cobranza judicial.”

Créditos Atrasados (criterio SBS) / Créditos Directos (%): Según la SBS (2015) se define: “Porcentaje de los créditos directos que se encuentra en situación de vencido o en cobranza judicial. El criterio de la SBS para considerar un crédito en situación de vencido, depende del número de días de atraso según el tipo de crédito: para los créditos corporativos, a grandes y a **medianas empresas cuando el atraso supera los 15 días; para los créditos a pequeñas y microempresas los 30 días;** y para los créditos hipotecarios y de consumo, a los 30 días de atraso se considera la cuota como vencida y a los 90 días de atraso el saldo total.”

Cobertura - Provisiones / Créditos Atrasados (%): Según la SBS (2015) se define: “Porcentaje de los créditos directos en situación de vencido o en cobranza judicial que se encuentran cubiertos por provisiones.”

La SBS a julio de 2018 ha reportado los siguientes indicadores a nivel de sistema financiero y retirando del análisis al banco de la Nación y Agrobanco:

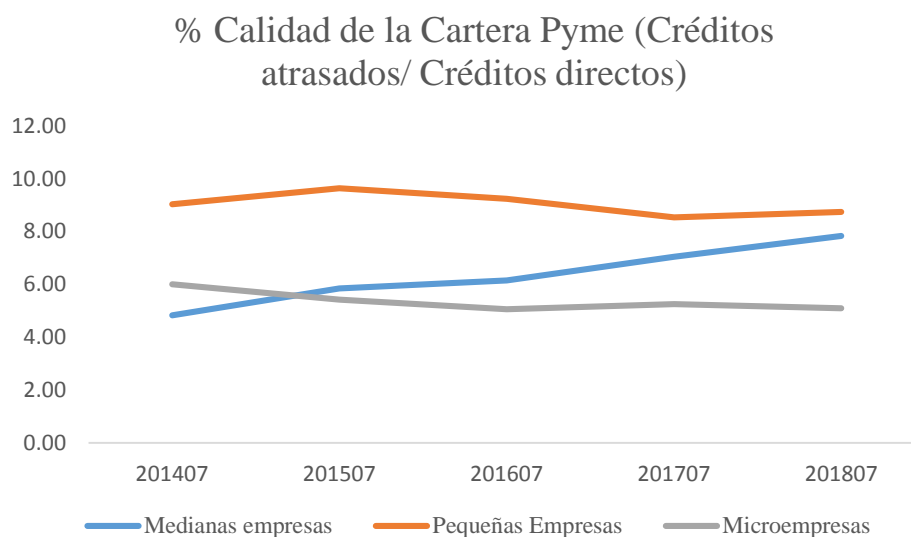
Figura IV.1 Créditos Directos Pyme



Fuente: SBS (2018) y Elaboración propia

En la figura IV.1, se observa que existe crecimiento sostenido desde el año 2014 en las colocaciones de microempresas, pequeñas empresas y medianas empresas.

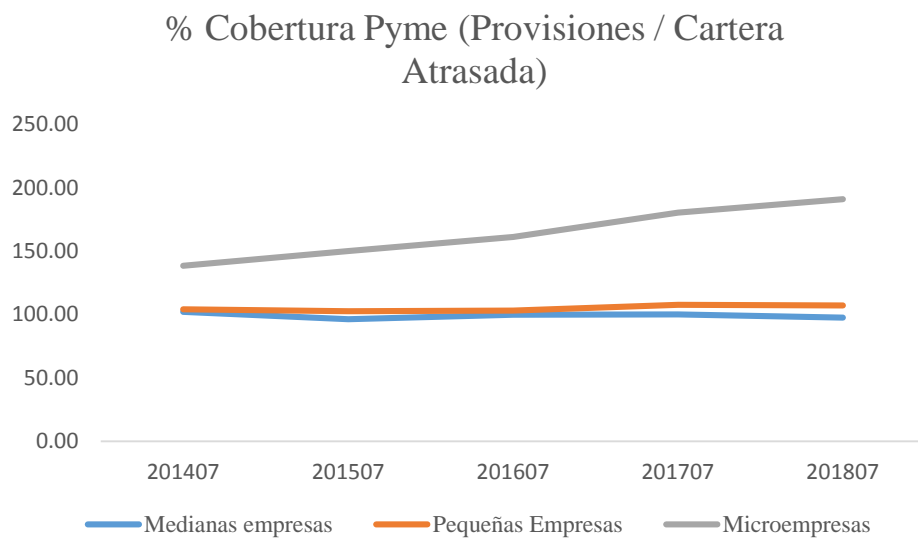
Figura IV.2 Calidad de la Cartera Pyme



Fuente: SBS (2018) y Elaboración propia

En la figura IV.2, se aprecia un deterioro en la calidad de la cartera para pequeñas empresas y medianas empresas.

Figura IV.3 Cobertura Pyme



Fuente: SBS (2018) y Elaboración propia

En la figura IV.3, se aprecia que el nivel de cobertura para microempresa se incrementa desde el 2014 debido a la estabilidad de la cartera atrasada y a un constante incremento de provisiones.

CAPÍTULO IV. MODELOS Y METODOLOGÍAS DEL CREDIT SCORING

En este capítulo se presenta las funciones de enlace para el Credit Scoring, justificaremos porqué se escoge a la metodología de regresión logística.

Abdou & Point (2011); Glennon, Kiefer, Larson, & Choi (2008); Saavedra-García & Saavedra-Garcia (2010), muestran los múltiples modelos matemáticos y econométricos utilizados en el Credit Scoring.

Rodriguez-Guevara, D. E., Becerra-Arévalo, J.A y Cardona-Valencia, D. (2017). Modelos y metodologías de Credit score para personas naturales: una revisión literaria. Revista CEA, 3(5), 13-28. Establecen que los modelos más representativos de Credit score hasta el año 2015 se parten en tres tipos de modelos generales: modelos paramétricos, modelos no paramétricos y modelos semi paramétricos.

4.1. Modelos Paramétricos Multivariantes

Estos modelos tienen la condición matemática de describir la sensibilidad de discriminación por medio de parámetros o estimaciones obtenidas bajo el proceso MCO (Mínimos Cuadrados Ordinarios).

Rodríguez-Guevara, D. E., Becerra-Arévalo, J.A y Cardona-Valencia, D. (2017) mencionan que:

- **Análisis discriminante**

Fisher (1936), Puertas & Martí (2012) y Mures, García, & Vallejo (2011), explican que el análisis discriminante es una relación de comportamiento de múltiples variables identificadas frente a una variable dependiente, la cual estructura varios posibles resultados o grupos de ecuaciones lineales que identifican la combinación eficiente para separar o discriminar a los buenos o malos pagadores.

El proceso distingue una serie de modelos lineales tal que así (1):

$$X = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_i x_i \quad (1)$$

Dónde:

λ_i = Representa los parámetros de la relación de combinaciones de variables.

X , representa una cantidad n de modelos posibles que nacen de las combinaciones entre variables. Cuando se tuviesen los parámetros hallados por cada grupo de datos, se toma la serie de datos en diferencia entre las medias (desviaciones estándar) quedando (2):

$$D = \lambda_1 d_1 + \lambda_2 d_2 + \dots + \lambda_i d_i \quad (2)$$

De esto, se busca encontrar una función de varianzas y covarianzas que determine el modelo con menos valores residuales entre ellos.

$$S = \sum_{p=1}^n \sum_{q=1}^n \lambda_p \lambda_q S_{pq}$$

Cuando se obtienen las estimaciones, la predicción depende de la puntuación de corte óptimo que discrimina los valores 0 y 1 de forma perfecta.

- **Modelos LDA (Linear Discriminant Analysis) - Modelo Z - Score**

De la misma forma que Fisher (1936) obtuvo el proceso discriminante, estos modelos fueron usados por Altman (1968), Altman (1980) y Elliott, Siu & Fung (2014) en sus estudios para identificar en el primer caso la probabilidad de quiebra de una entidad, que después utilizó para estudiar la probabilidad de impagos en una entidad bancaria, a dichos trabajos se les conoció como modelos LDA, Modelos Z y modelos Z – score; la ecuación del modelo Z es (3):

$$Z = \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (3)$$

Donde:

Z = Índice general

β_k = Son los parámetros de cambio que afectan la probabilidad de impago

X_{ik} = Variables cualitativas o cuantitativas.

- **Modelos Least – Absolute - Value (LAV)**

Glorfeld (1990) demuestra que el modelo lineal discriminante tiene falla, debido a la violación de los supuestos lineales; este autor propone hacer un análisis de Credit score tomando un modelo OLS (Ordinary Least Squares), condicionando el vector de residuales contrastados contra el valor absoluto de la diferencia entre los datos reales y los pronosticados, siendo este el determinante del mínimo error expresado.

Para ello, es requerido minimizar la sumatoria de los errores tal que (4):

$$\min z = \sum_i^n (\varepsilon_i^+ - \varepsilon_i^-), i = n \quad (4)$$

Y, dado que el modelo lineal requiere un valor cercano a cero, es imprescindible que dichos errores sean totalmente positivos (5).

$$y_i - (\alpha + \sum_j^n \beta_j x_{ij}) + \varepsilon_i^+ - \varepsilon_i^- = 0 \quad (5)$$

Donde:

y_i = Es la respuesta bivariada de riesgo de impago.

α, β_j = son los parámetros de cambio que afectan la probabilidad de impago.

X_{ik} = Variables cualitativas o cuantitativas.

ε_i = Residuales de la función

- **Modelos Lineales Probabilísticos (MLP)**

Como los describen Hardy & John (1985), Santos & Famá (2007), Bumacov, Ashta & Singh (2014), Puertas & Marti (2012), los modelos lineales probabilísticos son modelos en donde la mejor combinación lineal de las variables en estudio puede arrojar una respuesta bivariada eficiente, pero son sensibles a presentar problemas de especificación si violasen los supuestos de linealidad (Gujarati, 2004) (6). Son dos las desventajas los estimadores ineficientes y la probabilidad estimada puede estar fuera del intervalo [0,1] (Kim 2005).

$$Y_i = \alpha + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (6)$$

Donde:

Y_i = es la respuesta bivariada de riesgo de impago.

α, β_k = son los parámetros de cambio que afectan la probabilidad de impago.

X_{ik} = Variables cualitativas o cuantitativas.

- **Modelo Logístico (Logit)**

Según, Constancioara (2011); Gonçalves & Braga (2008); Lipovetsky & Conklin (2004); Majer (2006); Rayo, Lara, & Camino (2010); Rodríguez & Trespalacios (2015); Támara, Aristizábal, & Velásquez (2010), el modelo logístico es una respuesta para el incorrecto funcionamiento de los modelos MLP, este usa la función logística matemática para determinar un crecimiento condicionado en donde los valores más cercanos a cero (incluyendo $-\infty$) sean iguales a cero, y a medida de crecen al $+\infty$, tomarán la forma del valor 1, por lo tanto son modelos con parámetros no lineales que nacen de la máxima verosimilitud.

Su funcionamiento como un modelo econométrico dependerá del uso de la bondad de ajuste, matriz de error y curvas ROC para determinar el nivel de discriminación propuesto (7). Esta técnica es parte de los modelos machine learning.

$$z_t = \alpha + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik} \quad (7)$$

Siendo el modelo en la función logística (8):

$$P_i = E(y = 1|X) = \frac{1}{1+e^{-(\alpha+\beta_2 X_{t2}+\dots+\beta_k X_{tk})}} + \varepsilon_i = \frac{1}{1+e^{-z_t}} + \varepsilon_i \quad (8)$$

Donde:

Y_i = es la respuesta bivariada de riesgo de impago.

α, β_k = son los parámetros de cambio que afectan la probabilidad de impago.

X_{ik} = Variables cualitativas o cuantitativas.

- **Modelo Probabilístico (Probit)**

Así mismo, Melo & Granados (2011); S. Moreno (2013); Rayo et al. (2010); Roszbach (2004); Támara et al. (2010) y Webster (2011) muestran que los modelos Probit también han sido ampliamente usados por su condición de establecer normalidad al proceso de predicción, dándoles a las variables mayor estabilidad en el proceso de ajuste.

Su uso dependerá entonces de la función normal (9):

$$P_i = P(Y = 1|X) = P(I_t^* \leq I_t), \text{ que es igual a}$$

$$P_i = P(Z_t \leq \alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik}), \text{ igual a}$$

$$P_i = F(\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik})$$

Se dará entonces la CDF de la distribución normal para el modelo lineal, quedando (10):

$$F(I) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha+\beta_2 X_{i2}+\dots+\beta_k X_{ik}} e^{-\frac{z^2}{2}} dz + \varepsilon_i \quad (10)$$

Donde:

Y_i = es la respuesta bivariada de riesgo de impago.

α, β_k = son los parámetros de cambio que afectan la probabilidad de impago.

X_{ik} = Variables cualitativas o cuantitativas.

z = es la distribución normal estandarizada

- **Modelo Tobit**

Autores como, Fernández & Pérez (2005); Olagunji & Ajiboye (2010), Tan, Yen & Loke (2011) y Roszbach (2004) indican que dichos modelos proporcionan un mejor ajuste de la realidad de los datos al tener comprometida la información cuando existen con problemas de especificación de normalidad, condicionando la variabilidad de los errores. Si bien, estos modelos son similares al modelo Probit, su uso es menos comercial por su complejidad en el uso y que no siempre son tan generales para la construcción de un modelo de predicción de riesgo de crédito.

$$y_i^* = \alpha + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_k X_{ik} \quad (11)$$

Siendo y_i^* una variable no observable, la cual puede tener dos resultados

$$y_i = y_i^*, \text{ si } y_i > 0$$

$$y_i = 0, \text{ si } y_i \leq 0$$

Para todos los casos la función para obtener el resultado óptimo sería (12):

$$\ln L = \sum_{y_i > 0} [-\ln \sigma + \ln \phi(y_i - \alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik})] \sum_{y_i = 0} \ln \left[1 - \phi \left(\frac{\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik}}{\sigma} \right) \right] \quad (12)$$

Donde:

Y_i = es la respuesta bivariada de riesgo de impago.

α, β_k = son los parámetros de cambio que afectan la probabilidad de impago.

X_{ik} = Variables cualitativas o cuantitativas

ϕ y Φ = son funciones densidad acumulada para una distribución normal estándar

- **Modelos Logit Multinomiales**

Una propuesta alternativa al modelo Logit y Probit, Chaudhuri & Cheral (2012) y Gonçalves & Braga (2008) tienen en cuenta el uso que propone Basilea II en la forma que se debe analizar los créditos en rangos de maduración (A,B,C,D,E, siendo A, un buen pagador, y B,C,D,E pagadores morosos), por lo tanto, el cambio dirigido para los autores radica en que la variable Y, puede tener 5 respuestas posibles.

$P_i = Prob(Y_i=j)$, que es igual a (13):

$$P_i = \frac{e^{(\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik})}}{\sum_{j=0}^j e^{(\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik})}}, j \quad (13)$$

$$= 0, 1, 2, 3, 4$$

Donde:

Y_i = es la respuesta bivariada de riesgo de impago.

α, β_k = son los parámetros de cambio que afectan la probabilidad de impago.

X_{ik} = Variables cualitativas o cuantitativas

- **Modelos Logit Mixtos**

Kukuk & Rönnberg (2013) y Moreno (2013), propone una alternativa para los modelos Logit y Probit convencionales, esta autora recomienda su uso por su flexibilidad a la hora de predecir el nivel de crédito sin caer en los errores convencionales de los modelos Logit y Probit, ella muestra que usando un condicional binomial se obtendrá la siguiente ecuación (14), donde:

Donde:

Y_i = es la respuesta bivariada de riesgo de impago.

α, β_k = son los parámetros de cambio que afectan la probabilidad de impago.

X_{ik} = Variables cualitativas o cuantitativas

$\tilde{\beta}, \tilde{\theta}$ = parámetros hallados por máxima verosimilitud

$P(y_i = 1|x_i, \tilde{\beta}, \tilde{\theta}) = \int_{\beta_i} \Lambda(\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik}) f(\beta_i|\tilde{\beta}, \tilde{\theta}) d\beta_i$, que es igual a:

$$P(y_i = 1|x_i, \beta_k) = \Lambda(\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik}) = \frac{e^{(\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik})}}{\sum e^{(\alpha + \beta_2 X_{i2} + \dots + \beta_k X_{ik})}} \quad (14)$$

La diferencia encontrada frente al Logit convencional, es que los modelos permiten utilizar parámetros simulados, cuando se establece esta simulación, se pone también a

favor el hecho que los residuos son simulados, estandarizándolos y creando una mejor predicción.

4.2. Modelos No Paramétricos

A diferencia de los modelos paramétricos, los modelos no paramétricos se estructuran en procesos matemáticos que ocultan el proceso interno y se especifican en las variables de entrada y la de salida; son normalmente usados con procesos de nodos o redes que asemejan al cerebro.

Rodríguez-Guevara, D. E., Becerra-Arévalo, J.A y Cardona-Valencia, D. (2017) mencionan que:

- **Redes Neuronales (Neuronal networks)**

Desai, Crook & Overstreet (1996); West (2000); Esteve (2007); Pérez & Fernández (2007) y Soydaner & Kocadağlı (2015), explican que el uso de las redes neuronales utilizan un sistema artificial que se asemeja al cerebro humano, y que es capaz de identificar comportamientos variables de varios individuos a la vez dando un resultado de predicción eficiente, su estructura se basa en neuronas de información que se interconectan y determinan bajo entrenamiento de información, el resultado más similar a la realidad. Para su desarrollo es requerido el uso de un perceptron multicapa, que se estimula con ecuaciones que definen una entrada y una salida y comportamiento de los datos a modo de entrenamiento; cuando esté totalmente entrenado es capaz de re direccionar nuevos datos a resultados predecibles.

- **SVM (Support - Vector - Machine)**

Martens et al. (2010); J. F. Moreno & Melo (2011) y Zhou, Lai & Yen (2009) identifican un proceso nacido de las redes neuronales que usa geometría euclidiana para discriminar correctamente los datos; la operación matemática se asemeja a una operación lineal que identifica los espacios que existen entre los datos, son llamados «hiperplanos», que muestran la distancia discriminante que hay en respuestas de tipo binomial [0,1]. El factor de entrenamiento de los datos sería (15):

$$y(x) = \text{sign}(w^T x + b) \quad (15)$$

Que identifican dos posibles relaciones de discriminación (16):

$$\begin{cases} w^T x + b \geq +1, \text{ si } y_k = +1 \\ w^T x + b \leq -1, \text{ si } y_k = -1 \end{cases} \quad (16)$$

Donde:

y_k = es un valor -1 o 1 que determina la clase a la que pertenece x , siendo este un vector real de carácter p dimensional.

w = es un vector normal en el hiperplano

Siendo esta una función lineal clásica que tomará el nombre de hiperplano cuya función es analizar las distancias de cada punto y se bifurca en posiciones de $[-1, +1]$ mostrando la ecuación óptima que sirve para entrenar y discriminar la información, proporcionando una predicción óptima.

- **Modelos Bayesianos (Naive Models)**

Chang, Fung, Lucas, Oliver & Shikaloff (2000), Baesens, Castelo & Vanthienen (2002), Mileris (2010) y Webster (2011) usan modelos de predicción netamente probabilísticos que usan una función tal que (17):

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (17)$$

Donde $P(C_i|X)$ es la probabilidad posterior condicionada de C_i en X ; $P(C_i)$ es la probabilidad de C_i ; $P(X|C_i)$ es la probabilidad de X condicionada en C_i ; $P(X)$ es la probabilidad de X ; siendo esta una función maximizada para todos los casos de los clientes, sobre esta función es requerido hacer un entrenamiento de los datos quedando que (18):

$$P(X|C_i) = \prod_{k=1}^n P(x_1|C_i) * P(x_2|C_i) * ... * P(x_n|C_i) \quad (18)$$

Esta probabilidad será entonces una función normalizada para los errores que puedan proporcionar. Esta metodología trabaja esencialmente como un árbol de decisión, y puede ser amplificada si se tiene una variable de respuesta (A, B, C, D, E) mostrando una flexibilidad mayor en su uso.

4.3. Modelos Semi-Paramétricos (Hybrid Models)

Abdou & Pointon (2011); Akkoç (2012); Mileris (2010) y Moreno & Melo (2011) identifican que los modelos paramétricos pueden tener problemas de especificación a la hora de verificar la eficiencia de los errores de los modelos; si bien pueden tener errores normales, pueden no ser del todo eficientes e insesgados, lo que produce problemas de especificidad.

Teniendo en cuenta lo anterior, el uso de modelos que involucren procesos de redes neuronales o de análisis genético o discriminador por medio de máquinas vectoriales potencializan los procesos de identificación de las variables eliminando el proceso de sesgamiento de los datos mostrando una serie de datos normalizada y eficiente, prometiendo que un uso de cualquier modelo paramétrico ofrezca la identificación de un modelo no solo insesgado, también altamente eficiente.

Según Rodríguez-Guevara, D. E., Becerra-Arévalo, J.A y Cardona-Valencia, D. (2017) mencionan que:

- **Maximum Score**

Bult (1993) hace un comparativo de los modelos semi-paramétricos contrastados con un modelo paramétrico estándar como lo es el modelo Logit; según el autor, la desventaja del uso de los modelos paramétricos se enfoca en que su estimación no proporciona un dato robusto por su amplio nivel de supuestos implícitos. Los modelos semi-paramétricos en cambio combinan el proceso de los estimadores con funciones no

paramétricas que identifican una función de probabilidad de densidad normal para los errores. Para ello, la condición ecuacional será (19):

$$y_i^* = \begin{cases} 1, & \text{si } y_i = \beta'x_i + u_i > 0 \\ 0, & \text{en otro caso} \end{cases} \quad (19)$$

Donde y_i es una variable respuesta inobservable y y_i^* es el indicador de respuesta; cuando se obtiene dicho valor inobservable, se someterá a un proceso probabilístico tal que (20):

$$Prob(y_i^* = 1 | x_i) = Prob(y_i > 0 | x_i) = Prob(\beta'x_i + u_i > 0 | x_i) = Prob(u_i > -\beta'x_i | x_i) = 1 - F(-\beta'x_i)$$

$$S(\beta) = \sum_{i=1}^N S_i(\beta) = \sum_{i=1}^N y_i^* \text{sign}(\beta'x_i) \quad (20)$$

Donde:

Y_i = es la respuesta bivariada de riesgo de impago.

α, β_k = son los parámetros de cambio que afectan la probabilidad de impago.

X_{ik} = Variables cualitativas o cuantitativas

- **Técnica CHAID (CHI-squared Automatic Interaction Detection)**

Antipov & Pokryshevskaya (2010), Espin-García & Rodríguez-Caballero (2013) explican en su tesis que los árboles de decisión son usados para optimizar el proceso que una regresión logística o paramétrica no puede contrastar fácilmente, y es que los errores presentados en las regresiones paramétricas presentan fallas demasiado amplias cuando la base de datos no es eficiente. Los autores manifiestan que el uso de árboles de decisión antes de operar los modelos paramétricos ofrece un uso prudente de las variables antes de operarlas, evidenciando cuáles serán los parámetros más eficientes e insesgados antes de cualquier operación econométrica.

- **LS-SVM (Least Squares – Support Vector Machines)**

Un uso diferenciado mostrado por Tsai (2008), Zhou et al. (2009) del uso de un modelo mezclado entre las máquinas de soporte vectorial y los modelos lineales, estos procesos primero pueden discriminar la información de los datos por medio de un «kernel», que es un centro neurálgico de entrenamiento de los datos que permite identificar de forma no lineal el mejor proceso de discriminación, y sobre este determinar un modelo lineal que permite la relación matemática para un pronóstico eficiente de los datos. Si bien estos modelos no son del todo paramétricos y difíciles en algunos casos de crear, son muy potentes y tienen una aceptabilidad grande al momento de dar un resultado matemático.

- **Neuronal Network + Logit Models**

Sustersic, Mramor & Zupan (2007) presentan un uso sofisticado para los modelos logísticos muy similar al usado en la técnica CHAID, dando uso de redes neuronales para identificar las variables optimas usando un sistema de PCA (Análisis de Componentes Principales), acto seguido utiliza un proceso de modelos logísticos para identificar el modelo econométrico que permite estimar el discriminador de Credit Score más apropiado.

- **Decision Tree – CART (Clasification and Regression trees) models**

Zhang, Zhou, Leung & Zheng (2010), Baklouti (2014); Díaz Sepulveda (2012); Kočenda & Vojtek (2009) explican que los árboles de decisión basados en funciones bayesianas pueden ser usados como modelos de discriminación para el Credit score, los árboles de decisión se comportan de alguna manera similar una red neuronal dirigida, es necesaria una técnica más sofisticada para no dejar en paralelo el hecho que los árboles de decisión solamente por medio de los nodos o variables no son capaces de dar un valor único de salida, por lo que el uso de regresión lineal es necesaria en dichos procedimientos (21).

$$i_t = \phi(\{Y = 1 \setminus \tau\}) \quad (21)$$

Donde τ es el nodo y ϕ define la probabilidad de éxito del nodo. Para la relación lineal de esta expresión se encuentra entonces (22):

$$i(\tau) = \sum_{\text{suje } i \in \tau} (Y_i - \bar{Y}(\tau))^2 \quad (22)$$

Donde Y_i será (23):

$$Y = F(x_1, x_2, x_3, \dots, x_p) = \alpha + \sum_{j=1}^p \beta_k x_j = \alpha + \sum_{j=1}^p \beta_k g_j(x) = f(x) \quad (23)$$

Donde:

Y_i = es la respuesta bivariada de riesgo de impago.

α, β_k = son los parámetros de cambio que afectan la probabilidad de impago.

X_{ik} = Variables cualitativas o cuantitativas

4.4. Ventajas y desventajas de los Modelos Paramétricos, No Paramétricos y Semi Paramétricos

Se presenta la tabla III.1 se resumen de las ventajas y desventajas de los modelos paramétricos, no paramétricos y semi paramétricos.

Tabla III.1 Ventajas y desventajas de los modelos paramétricos, no paramétricos y semi paramétricos

	Modelos	Ventajas	Desventajas
Paramétricos	Análisis discriminante	▪ Buen rendimiento grandes muestras	▪ Estimadores ineficientes
	Modelos LDA (Linear Discriminant Analysis) Modelo Z-Score	▪ Combinación lineal de ratios financieros para predecir quiebras de empresas	▪ Solo aplica para empresas y no en personas.
	Modelos Least- Absolute – Value (LAV)	▪ Minimiza la suma de valores absolutos de los residuos, estimación robusta.	▪ Similar a los mínimos cuadrados ordinarios, presenta múltiples soluciones y son inestables cuando se presentan valores atípicos.
	Modelos Lineales Probabilísticos (MLP)	▪ Si es una Combinación lineal de variables optimas puede predecir una respuesta bivariada eficiente.	▪ Pierde precisión si se violan los supuestos de linealidad.
	Modelo Logístico (Logit)	▪ Gran rendimiento, buenas propiedades estadísticas. Corrige la limitaciones de MLP	▪ Fácil interpretación de la data.
	Modelo Probabilístico (Probit)	▪ Buen rendimiento, normalidad en la predicción, mayor estabilidad de variables.	▪ Difícil interpretación de los datos
	Modelos Tobit	▪ Proporcionan un mejor ajuste a la realidad de los datos	▪ Su uso es menos comercial por su nivel de complejidad

No Paramétricos	Modelos logit multinomiales	<ul style="list-style-type: none"> Generaliza la regresión logística con más de dos posibles resultados discretos. 	<ul style="list-style-type: none"> El aprendizaje es lento si las categorías a predecir son muchas, hay algoritmos que aprenden más rápido como Naive Bayes.
	Modelos Logit mixtos	<ul style="list-style-type: none"> No tiene las limitaciones del logit estándar, variaciones estándar de la muestra. Permite utilizar parámetros simulados. 	<ul style="list-style-type: none"> Estimación compleja de parámetros mediante simulación de Monte Carlo y genera mayor error tipo I.
	Redes Neuronales (Neural networks)	<ul style="list-style-type: none"> Gran predicción, flexibilidad 	<ul style="list-style-type: none"> Difícil interpretación.
	SMV (Support – Vector - Machine)	<ul style="list-style-type: none"> Gran predicción 	<ul style="list-style-type: none"> Difícil interpretación
Semi Paramétricos	Modelos Bayesianos (Naive Models)	<ul style="list-style-type: none"> Requiere una muestra pequeña de datos 	<ul style="list-style-type: none"> Se limita a las características que ingresan al modelo.
	Máximo Score	<ul style="list-style-type: none"> Obtiene estimadores robustos de los parámetros del modelo 	<ul style="list-style-type: none"> Problemas de heterocedasticidad y correlación.
	Técnica Chaid (Chi - squared automatic interaction detection)	<ul style="list-style-type: none"> Los resultados son altamente visuales y de fácil interpretación. 	<ul style="list-style-type: none"> Necesita tamaños de muestra bastante grandes por la cantidad de divisiones que realiza.
	LS-SVM (Least squares-support vector machines)	<ul style="list-style-type: none"> Utiliza Kernel o núcleos y son estadísticamente bien fundados. 	<ul style="list-style-type: none"> Difícil interpretación
	Neuronal Network + Logit Models	<ul style="list-style-type: none"> Se aplica varias técnicas como componentes principales, algoritmo genético en la red neuronal de kohonen. 	<ul style="list-style-type: none"> Elevado esfuerzo computacional.
	Decision tree (Clasificación and regression trees)	<ul style="list-style-type: none"> Fácil de entender, requiere poca preparación de datos, visual, eficaz para predecir y clasificar. 	<ul style="list-style-type: none"> Sobreajuste y complejidad computacional.

Fuente: Capítulo IV y Elaboración propia

4.5. Elección de función de enlace

Rodríguez Guevara, D., Becerra Arévalo, J. & Cardona Valencia, D. (2017), destacan el uso de los modelos logit, los modelos LDA, los modelos de regresión censurada, los modelos LS-SVM, SVM, las redes neuronales, los arboles de decisión CART y los modelos bayesianos.

No existe un modelo o método perfecto de identificación de Credit Score por varias situaciones, la primera de ellas es la calidad de los datos y las variables presentadas; la dependencia o el uso de un modelo dependerá de la calidad de la información y sobre que variables se deberán trabajar; pero tomando un punto estricto en la revisión bibliográfica, el modelo más utilizado por los autores es el modelo logit o probit, por su facilidad de interpretación, su facilidad de manejo en proceso matemático, y si bien, no son los más óptimos si son un referente base para la discriminación de clientes Guevara et al., (2017)

En el presente trabajo la función de enlace a utilizar es la regresión logística binaria o logit, porque dicha técnica estadística es una herramienta intuitiva, fácil de explicar, su mantenimiento es sencillo, fácil de calibrar, rápida ejecución computacional, fácil comprensión y es aplicada por la mayoría de bancos a nivel mundial. El comité de Basilea sugiere el uso de modelos estadísticos entre ellos la regresión logística para la medición del riesgo de crédito. El sistema financiero peruano utiliza la regresión logística ya que es una técnica aceptada y evaluada por la SBS en sus visitas regulatorias a las instituciones financieras del país.

4.6. Análisis de la función de enlace

Torosyan (2017) la regresión logística binaria es un análisis de regresión donde la variable dependiente es binaria, contiene datos codificados como 1 y 0. La regresión logística binaria es un análisis predictivo.

El objetivo de la regresión logística binaria es encontrar el modelo, que describe la relación entre características de interés (variable dependiente) y el conjunto de variables independientes. Para entender el modelo general de regresión logística se define el odds.

Odds de un evento es la relación entre la probabilidad de que ocurra un evento y la probabilidad de que no ocurra.

Si la probabilidad de presencia de la característica de interés es p , la probabilidad de ausencia de la característica de interés es $1-p$. Entonces las probabilidades correspondientes son un valor dado por esta fórmula:

$$odds = \frac{p}{1-p} = \frac{\text{probabilidad de presencia de la característica}}{\text{probabilidad de ausencia de la característica}}$$

Como la regresión logística calcula la probabilidad de que un evento ocurra sobre la probabilidad de que un evento no ocurra, la influencia de las variables independientes generalmente se explica en términos de probabilidades.

Con la regresión logística, la media de la variable dependiente p en términos de la variable independiente x viene dada por la ecuación $p = \alpha + \beta x$. Esta ecuación no genera un buen modelo ya que los valores de $p = \alpha + \beta x$ no se encuentran entre 0 y 1. La regresión logística brinda una solución a este problema al transformar los odds usando el logaritmo natural.

Con la regresión logística se modela el logaritmo natural de odds, como una función lineal de variable independiente.

$$\text{logit}(y) = \ln(odds) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta x \dots (1)$$

Donde p es la probabilidad del resultado interesado y x es la variable independiente, los parámetros de la regresión logística son α y β . Este es un modelo simple de regresión logística. De la ecuación (1) podemos derivar una ecuación para calcular la probabilidad de predicción:

$$p = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} = \frac{1}{1+e^{-(\alpha+\beta x)}} \dots (2)$$

Suponiendo que tenemos predictores múltiples, podemos construir un modelo logístico general como:

$$\text{logit}(y) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k \quad \dots (3)$$

De la ecuación (3) nosotros podemos derivar una ecuación para predecir la probabilidad:

$$p = \frac{e^{\alpha + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\alpha + \beta_1 x_1 + \dots + \beta_k x_k}} = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \dots + \beta_k x_k)}} \quad \dots (4)$$

4.7. Curva logística

Torosyan (2017) menciona que: Cuando la variable dependiente es binaria y la variable independiente es numérica, el modelo logístico ajusta a una curva a la relación entre x e y . La curva logística es una forma común “S” (curva sigmoidea).

Una función logística simple se define mediante la siguiente formula:

$$y = \frac{e^x}{1 + e^x}$$

Esta ecuación puede ser extendida de la forma:

$$y = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} = \frac{1}{1 + e^{-(\alpha + \beta x)}}$$

Expandiendo matemáticamente (Kleinbaum y Klein, 2002):

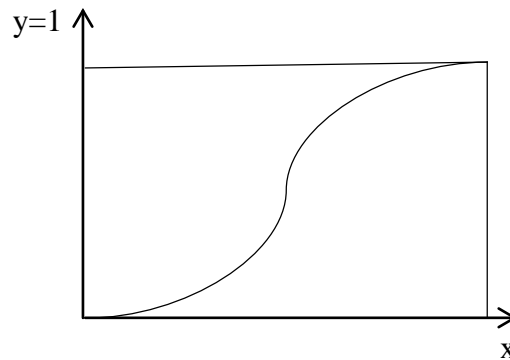
$$y = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x_1 - \dots - \beta_n x_n))}$$

dónde:

- ✓ β_0 es el término independiente
- ✓ β_1, \dots, β_n son los pesos de las n variables consideradas
- ✓ x_1, \dots, x_n son las n variables consideradas
- ✓ y : es una probabilidad que toma valores 0 y 1.

En la figura III.1 se muestra la función logística como se observa el valor de “y” toma valores entre 0 y 1.

Figura III.1 Función logística



Fuente: Torosyan (2017) y Elaboración propia

4.8. Suposiciones de la regresión logística

Torosyan (2017) a diferencia de los modelos lineales generales, la regresión logística binaria no tiene muchas suposiciones clave, particularmente no requiere una relación lineal entre la variable dependiente e independiente, la normalidad de la distribución de errores, la homocedasticidad de los errores y el nivel de medición de las variables independientes.

Sin embargo, la regresión logística requiere de otras suposiciones:

1. La regresión logística binaria requiere que las variables dependientes sean binarias.
2. Dado que la regresión logística binaria asume que $P(Y)=1$ es la probabilidad que ocurra el evento, esto requiere que la variable dependiente sea codificada con valores 1 y 0.
3. El modelo debe estar ajustado correctamente, esto significa que todas las variables significativas deben ser incluidas. El modelo no debe estar sobrecargado con variables sin sentido de negocio.

4. La regresión logística binaria requiere que cada observación sea independiente. Además, debería tener poca o ninguna multicolinealidad, lo que significa que las variables independientes no son funciones lineales entre sí.
5. La regresión logística binaria requiere linealidad de la relación entre variables independientes y logaritmo odds. Mientras tanto no requiere una relación lineal entre variables dependientes e independientes.
6. La regresión logística binaria requiere tamaño de muestra bastante grande. Los estudios con tamaños de muestra pequeños sobreestiman los efectos de medida. Las variables más independientes se deben incluir en el modelo y es necesario que el tamaño de muestra sea elevado.

4.9. Estimación de máxima verosimilitud

Torosyan (2017) aunque la regresión logística se parece al modelo de regresión lineal simple, la distribución subyacente es binomial, y los parámetros α y β no se pueden estimar de la misma manera que la regresión lineal simple.

Los coeficientes generalmente se estiman mediante el modelo de máxima verosimilitud (Park, Hyeoun- Ae, April 2013) la verosimilitud es una probabilidad de obtener valores observados de la variable dependiente dados los valores observados de las variables independientes. La probabilidad varía entre 0 y 1 como cualquier otra probabilidad.

Se muestra:

$$P = (Y = y_i) = P_i^{1-y_i}(1 - P_i)^{y_i}$$

Donde P_i es la probabilidad de la i -ésima observación, y_i es el valor de la variable aleatoria Y que toma valor 0 o 1. Suponiendo que nuestra n observaciones son independientes la verosimilitud de los datos es igual a:

$$L = \prod_{i=1}^n P_i^{1-y_i}(1 - P_i)^{y_i}$$

El método de máxima verosimilitud proporcionara valores para α y β que maximizan la función L.

4.10. Evaluación del modelo de regresión logístico binario

Hay varias formas de estimar la regresión logística. En primer lugar, se debe evaluar el modelo general. En segundo lugar, es necesario evaluar la importancia de cada variable explicativa. En tercer lugar, se evalúa la precisión predictiva.

4.10.1 Evaluación general del modelo

a) Prueba de razón de verosimilitud

En la evaluación general del modelo, se observa cuán solida es la relación entre las variables independientes y la variable dependiente. Si la regresión logística con k variables independientes demuestra una mejora sobre el modelo sin variables independientes (modelo nulo), entonces proporciona un mejor ajuste a los datos (Park Hyeoun-Ae, abril 2013). Esto se realiza utilizando la prueba de razón de verosimilitud, que compara la probabilidad de los datos en el modelo completo con la probabilidad de que los datos estén bajo el modelo sin variables independientes. El ajuste general del modelo con k coeficientes se puede evaluar mediante la hipótesis nula:

$$H_0: \beta_0 = \beta_1 = \dots = \beta_K$$

$-2 \log$ la probabilidad del método nulo se compara con la probabilidad de $2 \log$ del modelo dado. La probabilidad del método nulo es la probabilidad de obtener la observación si las variables explicativas no tienen impacto en el resultado. La probabilidad del modelo dado es la probabilidad de obtener la observación si todas las variables explicativas se incluyen en el modelo.

La diferencia de estos indica un índice de bondad de ajuste G, χ^2 statistic, con k grados de libertad. Mide qué tan bien influyen las variables independientes en la variable dependiente.

$$G = \chi^2 = (-2 \log \text{likelihood of null mode}) - (-2 \log \text{likelihood of the given model})$$

Si el valor de p para la estadística de ajuste general del modelo es inferior a 0.05 entonces rechaza H_0 con la conclusión de que al menos una de las variables independientes tiene un impacto en el resultado o en la variable dependiente.

b) Prueba de bondad de ajuste Chi cuadrado

Torosyan (2017), menciona que es una prueba no paramétrica que se utiliza para descubrir como el valor observado de un evento dado es significativamente diferente del valor esperado. La hipótesis es la siguiente:

Hipótesis Nula: Supone que no hay una diferencia significativa entre el valor observado y el valor esperado.

Hipótesis alternativa: Supone que hay una diferencia significativa entre el valor observado y el esperado. Si el valor p es menor que el nivel de significancia, la hipótesis nula es rechazada. En la regresión lineal los residuos se definen como $y_i - \hat{y}_i$ donde y_i es el valor observado de la variable para el sujeto i -ésimo. Para la regresión logística donde y_i es 1 o 0, la predicción correspondiente del modelo es como

$$\hat{y}_i = \frac{\exp(\alpha + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}{1 + \exp(\alpha + \beta_1 x_{i1} + \dots + \beta_k x_{ik})}$$

La prueba chi-cuadrado se basa en los residuos $y_i - \hat{y}_i$. Un residuo estandarizado se define como:

$$r_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i(1 - \hat{y}_i)}}$$

El estadístico χ^2 chi-cuadrado se puede formar como:

$$\chi^2 = \sum_{i=1}^n r_i^2$$

Esta estadística sigue la distribución χ^2 chi-cuadrado con $n-(k+1)$ grados de libertad.

c) Test Hosmer – Lemeshow

Para Torosyan (2017) el test de Hosmer – Lemeshow mide que tan bueno es el modelo. La prueba evalúa si las tasas de eventos observados coinciden con las tasas de eventos esperados en los subgrupos de la población del modelo.

El test se implementa dividiendo las probabilidades pronosticadas en diez grupos iguales, de acuerdo con sus valores (deciles).

La hipótesis es la siguiente:

H_0 : Las tasas de eventos reales y pronosticados son similares en 10 deciles.

H_1 : No son lo mismo.

El valor de las estadísticas de prueba es:

$$\chi^2 = \sum_{g=1}^{10} \frac{(O_g - E_g)^2}{E_g}$$

Donde los O_g son los eventos observados, y E_g son los eventos esperados para el g-ésimo grupo decil de riesgo. La estadística de prueba sigue asintóticamente una distribución χ^2 con 8 (grupos – 2) grados de libertad.

Valores pequeños con un valor de p grande más cercano a 1 significa un buen ajuste de datos. Los valores grandes con $p < 0.05$ significan un mal ajuste a los datos.

4.10.2 Significancia estadística de los coeficientes de regresión individuales

Después de evaluar el modelo general, el siguiente paso es evaluar la importancia de cada variable independiente. El coeficiente de la i-ésima variable explicativa indica el cambio en las probabilidades de registro pronosticadas para un cambio de unidad en la i-ésima variable explicativa, cuando todas las demás variables explicativas permanecen sin cambios.

a. Prueba de razón de verosimilitud

Según Torosyan (2017) el test se usa para evaluar el ajuste general del modelo. También se usa para evaluar la significancia estadística de los predictores individuales.

La estadística de prueba se calcula como:

$$G = -2 \ln \frac{L_0}{L_1} = -2(\ln L_0 - \ln L_1)$$

Esta estadística se compara con la distribución χ^2 con 1 grado de libertad.

b. Estadística de Wald

La estadística de wald se usa para evaluar la importancia de los coeficientes individuales en un modelo dado (Bewick et., al 2005). La estadística es la relación entre el cuadrado del coeficiente de regresión y el cuadrado del error estándar del coeficiente. El cálculo es el siguiente:

$$W_j = \left(\frac{\text{coefficient}}{SE_{\text{coefficient}}} \right)^2$$

Cada estadística de wald se compara con una distribución de χ^2 con un grado de libertad.

4.10.3 Precisión predictiva y discriminación

a. Tabla de clasificación

La tabla de clasificación es una herramienta para evaluar la precisión predictiva del modelo de regresión logística. En la tabla III.2 se muestra la tabla de clasificación, los valores observados para el resultado dependiente y los valores predichos se clasifican de forma cruzada. Por ejemplo, si el valor de corte es 0.5, todos los valores pronosticados superiores a 0.5 se pueden clasificar como predictores de un evento, y todos los valores por debajo de 0.5 no predicen el evento.

Tabla III.2 Tabla de clasificación de muestra

Observado	Predicho	
	0	1
0	a	b
1	c	d

Fuente: Torosyan, Application of binary logistic regression in Credit Scoring (2017) y Elaboración propia

Donde a y d son el número de casos que se predicen correctamente, b y c son el número de casos que no se predicen correctamente.

Si vemos muchos recuentos en las celdas a y d, y pocos en las celdas b y c, podemos concluir que nuestro modelo tiene buen ajuste. Se utilizan dos indicadores para evaluar la precisión de una prueba que predice la sensibilidad y especificidad binarias de los resultados. La sensibilidad es la proporción de verdaderos positivos ($Y=1$). La especificidad es la proporción de negativos verdaderos ($Y=0$).

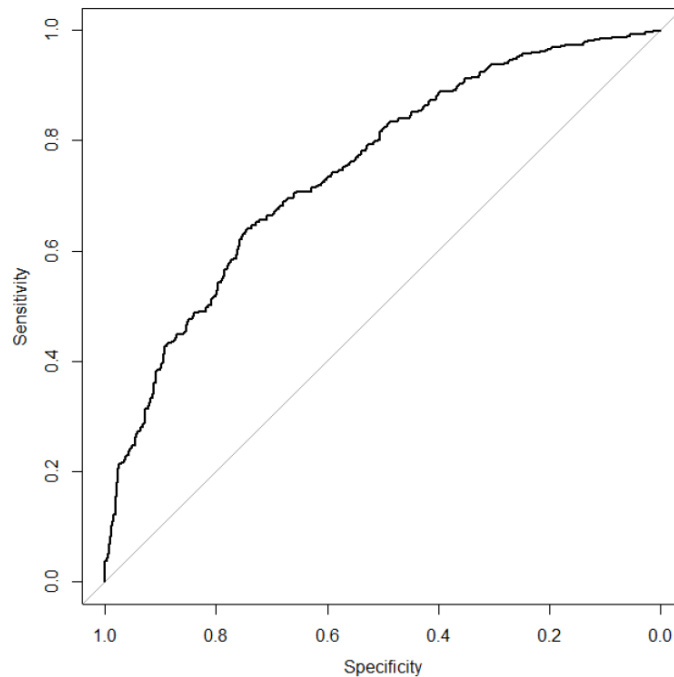
La sensibilidad se calcula con la fórmula $d/(c + d)$. La especificidad se calcula mediante la fórmula $a/(a + b)$. Los valores de sensibilidad y especificidad dependen del valor de corte que elijamos. Si aumentamos el punto de corte, se pronosticarán menos observaciones como positivas. En otras palabras, menos de $Y=1$ observaciones serán predichas como positivas, lo que significa que la sensibilidad disminuirá. Mientras que, más de las observaciones $Y=0$ serán predichas como negativas, lo que significa que la especificidad aumentará. El modelo es perfecto si tiene 100% de sensibilidad y 100% de especificidad. En la práctica, este resultado no suele ser alcanzable.

b. Curva ROC (Receiver Operating Characteristics)

Torosyan (2017) la curva ROC es una técnica fundamental para la evaluación de las pruebas de diagnóstico. Extiende la tabla de clasificación y muestra todos los puntos de corte entre 0 y 1. Para cada punto de corte posible, se puede formar una tabla de dos por dos. La curva ROC es un gráfico de los valores de sensibilidad versus uno menos la especificidad, ya que el valor de corte se incrementa de 0 a 1. Por ejemplo, en la

figura III.2 se aprecia la curva ROC, dicha curva es más informativa que la tabla de clasificación porque refleja el poder predictivo para todos los valores de corte.

Figura III.2 Ejemplo de la curva ROC



Fuente: Torosyan. Application of binary logistic regression in Credit Scoring (2017)

El área bajo la curva ROC es una forma de resumir la capacidad de discriminación de un modelo. Un área igual a 1 indica una discriminación perfecta. Mientras que un área de 0.5 indica una prueba sin valor. Un área de más de 0.7 se considera un modelo muy excelente.

4.11. Medición del buen desempeño de la regresión Logística

Indicador Kolmogorov - Smirnov (KS)

Mide el nivel predictivo del modelo. Identifica que tan separados están los clientes buenos de los clientes malos. Si el indicador es cercano a 100% el modelo es muy bueno.

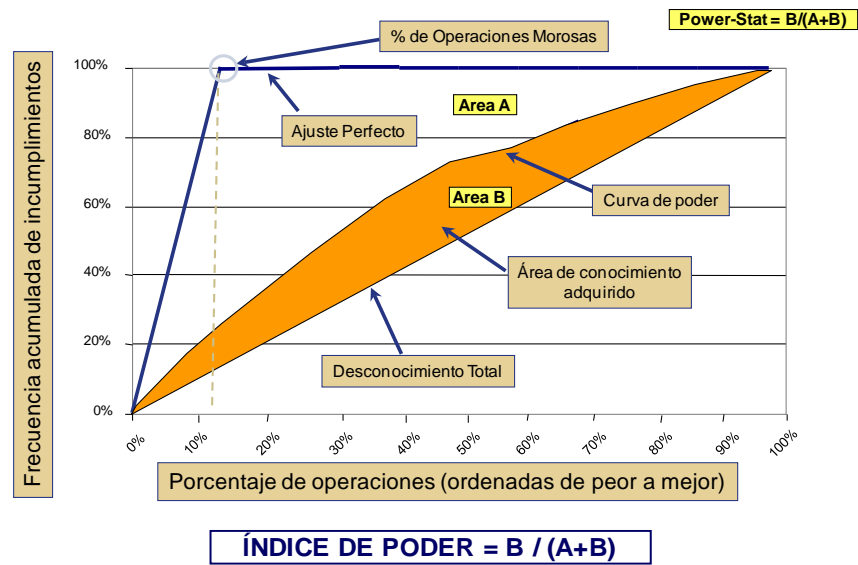
La fórmula es:

$$KS = | \% \text{ de malos clientes rechazados} - \% \text{ de buenos clientes rechazados} |$$

Indicador Gini:

El indicador Gini mide el poder predictivo del modelo basándose en la Curva de Poder Predictivo. En la figura III.3, se observa el área B que al dividirla entre el área A más el área B, nos da como resultado el valor del indicador Gini.

Figura III.3 Indicador Gini



Fuente: Capítulo IV y Elaboración propia

Mientras mayor sea el índice Gini, el modelo será más robusto.

CAPÍTULO V. CREDIT SCORING EN PYMES

Se realiza un diagnóstico sobre la aplicación del Credit Scoring en las Pymes y en las microempresas, se encuentra evidencia sobre la importancia de las variables del sistema financiero (RCC), la data del solicitante, por ejemplo, la zona, el sexo, la edad, el estado civil, la situación laboral, el local propio o alquilado, la raza del solicitante si es blanco o negro. La experiencia previa en la institución financiera, por ejemplo, la mora, las cuotas pagadas. El vínculo con el cliente se mide con la antigüedad en sus productos activos o pasivos, las garantías reales. Los activos y pasivos, el sector donde opera la Pyme ya sea agricultura, producción o servicios. En algunas aplicaciones toman el score de las centrales de riesgo y las combinan con las variables internas de la institución para predecir el incumplimiento. En los países donde la Pyme es formal se toma información de la situación financiera como el patrimonio, apalancamiento, capital de trabajo, etc. El modelo propuesto en la presente tesis recoge algunas variables revisadas en el presente diagnóstico y con ello propone un modelo de Credit Scoring versátil que pueda ser implementado en cualquier realidad de negocio ajustándose a los datos disponibles. A nivel predictivo cada uno de los modelos revisados están por encima de los estándares, sin embargo, la presente tesis propone un modelo aplicado a capital de trabajo revolvente y examina que el riesgo de la Pyme es el riesgo del titular.

La presente tesis parte de las experiencias revisadas y se enfoca en encontrar variables predictivas para explicar el incumplimiento y se sustenta en indicadores de performance como el coeficiente Gini y el ROC.

En este capítulo se presenta las aplicaciones del Credit Scoring para Pyme.

5.1. Aplicaciones del Credit Scoring

Se presenta las siguientes aplicaciones del Credit Scoring

- a. Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II

En la tabla V.1 Rayo et al. (2010) se propone las variables:

Tabla V.1 Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II

Fase del proceso de evaluación del microcrédito	Variable	Concepto	Signo esperado del estimador (β)
FASE 1	ZONA	Lugar geográfico de la agencia o sucursal. Variable dicotómica (0) Zona centro (1) Extra-radio	-
FASE 2	ANTIGUO	Tiempo del prestatario como cliente de la entidad. Variable numérica	-
	CRED_CONC	Créditos concedidos con anterioridad. Variable numérica	-
	CRED_CONC2	Créditos concedidos en el último año. Variable numérica	-
FASE 3A	CRED_DENEG	Créditos denegados con anterioridad. Variable numérica	+
	SECTOR	Sector de actividad de la microempresa. Variable categórica (0) Comercio (1) Agricultura (2) Producción (3) Servicio	+/-
FASE 3B	DEST_CRED	Destino del microcrédito. Variable dicotómica (0) Capital de trabajo (1) Activo Fijo	-
	CLAS_ENT	Clasificación del cliente según la IMF. Variable dicotómica (0) Cliente normal (1) Cliente con algún tipo de problema	-
	CUOT_TOT	Número total de cuotas pagadas en historial de crédito. Variable numérica	-
FASE 3C	CUOT_MORA	Número de cuotas incurridas en morosidad. Variable numérica	+
	MEDIA_MORA	Promedio (días) de la morosidad del cliente. Variable numérica	+
	MORA_MAYOR	Número de días de la mayor mora del cliente. Variable numérica	+
	SEXO	Género del prestatario. Variable dicotómica. (0) Hombre, (1) Mujer	+
	EDAD	Edad en el momento de la solicitud del crédito. Variable numérica	+
	E_CIVIL	Estado Civil. Variable dicotómica. (0) Soltero, (1) Unidad Familiar	+
	SIT_LAB	Situación laboral del cliente. Variable dicotómica (0) Propietario (1) Dependiente	-
	R1	Rotación Activos = Ingresos Ventas / Total Activo	-
	R2	Productividad = Utilidad Bruta / Costes Operativos	-
	R3	Liquidez = Capacidad Pago / Total Activo	-
	R4	Rotación Liquidez = Capacidad Pago / Ingresos Ventas x 360	+
	R5	Dependencia o Endeudamiento = Total Pasivo / (Total Pasivo + Total Patrimonio)	+
	R6	Apalancamiento= Total Pasivo/Total Patrimonio	+
	R7	ROA=Utilidad Neta/ Total Activo	-
	R8	ROE= Utilidad/Total Patrimonio	-

FASE 4	GARANT	Tipo de garantía aportada por el cliente. Variable dicotómica (0) Declaración jurada (1) Garantía Real (aval, prenda, hipoteca, etc)	-
FASE 5	MONEDA	Tipo de moneda en la concede el crédito. Variable dicotómica (0) Nuevos Soles (1) US\$	-
	MONTO	Importe del microcrédito. Variable Numérica.	-
	MONTO_RECH	Cantidad rechazada por la institución. Variable Numérica.	+
	DURACION	Número de cuotas mensuales del microcrédito solicitado. Variable numérica.	+
	INT_MENS	Tasa de interés mensual del microcrédito. Variable numérica.	+
	VTOCRED_SBS	Pronóstico del analista sobre la situación del crédito a su vencimiento. Variable dicotómica (0) Vigentes (1) Con problemas estimados.	-
FASE 6	PIB	Tasa de variación anualizada del Producto Interno Bruto durante la vigencia del crédito. Variable numérica.	-
	IPC	Tasa de variación anualizada del Índice de Precios al Consumidor durante la vigencia del crédito. Variable numérica.	+
	IE	Tasa de variación anualizada del índice de Empleo durante la vigencia del crédito. Variable numérica.	-
	TC	Tasa de variación anualizada de la Tasa de Cambio durante la vigencia del crédito. Variable numérica.	+
	TI	Tasa de variación anualizada de los tipos de Interés durante la vigencia del crédito. Variable numérica.	+
	IGB	Tasa de variación anualizada del Índice General Bursátil durante la vigencia del crédito. Variable numérica.	-
	AGUA	Tasa de variación anualizada de la tarifa municipal de Agua durante la vigencia del crédito. Variable numérica.	+
	LUZ	Tasa de variación anualizada de la tarifa municipal de Luz durante la vigencia del crédito. Variable numérica.	+
	TFNO	Tasa de variación anualizada de la tarifa de Teléfono durante la vigencia del crédito. Variable numérica.	+

Fuente: Rayo et al., Un modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II (2010) y Elaboración propia

b. Modelo Scoring para el otorgamiento de crédito de las Pymes

En la tabla V.2 (Valencia, 2017) propone en su tesis las siguientes variables:

Tabla V.2 Modelo Scoring para el otorgamiento de crédito de las Pymes

Variable	Concepto
No Pago	La variable toma valor 1 si los días de no pago es mayor o igual a 30 días y toma valor cero cuando paga menor a 30 días
Plazo	Es el plazo otorgado en la operación
Plazo2	Es el plazo al cuadrado
Cupo	Indica el monto total que es otorgado y la exposición global de riesgo
Ventas Netas	Es el rubro financiero que indica los ingresos que obtiene una empresa proveniente de la actividad que desarrolla.
Activo Total	Muestra el tamaño de una empresa.
Missing AP	Toma el valor 1 si no cuenta con un score de Acierta Plus
Prueba acida	Se define como: $(\text{Activo Corriente} - \text{Inventario}) / \text{Pasivo Corriente}$
Cubrimiento de intereses	Es una razón que indica el número de veces que una empresa puede pagar los intereses de la deuda con el Ebitda.
Razón Endeudamiento	Es una de las razones más fundamentales a la hora de asumir el riesgo con una Pyme, puesto que muestra cuántos de sus activos están financiados con deuda

Fuente: Valencia. Modelo Scoring para el otorgamiento de crédito de las Pyme (2017) y Elaboración propia

c. Score de Bureau para cajas municipales de ahorro y créditos

En la tabla V.3, los autores Hidalgo & Pichardo (2015) presentan las siguientes variables:

Tabla V.3 Score de Bureau para cajas municipales de ahorro y créditos

Variable	Concepto
MAXENTI2M	Máximo número de entidades donde el cliente registra deudas y obligaciones financieras en los últimos 12 meses. A mayor cantidad de entidades, mayor será su probabilidad de calificar como un cliente malo
MAXCLAS12M	Máxima clasificación en el sistema financiero que presenta un cliente los últimos 12 meses. A mayor clasificación de riesgo que presente el cliente, mayor será su probabilidad de riesgo de calificar como cliente malo.
DIFENT12M	Diferencia entre el número de entidades del último mes y hace 12 meses. A mayor diferencia de entidades, mayor será la probabilidad de que el cliente califique como malo.
RATIO_SALDOS12M	Ratio entre el ultimo saldo deudor reportado en los últimos 12 meses y el máximo saldo registrado en dicho periodo. Cuanto mayor sea el ratio, habrá mayor probabilidad de que el cliente califique como malo.
MAXUSOLC6M	Máximo uso de la línea de crédito en los últimos 6 meses medido en porcentaje. Si el cliente en los últimos 6 meses presenta un mayor ratio de uso de su línea de crédito, mayor será la probabilidad que califique como cliente malo
TCREDCON6M	El cliente registra créditos de consumo en los últimos seis meses. Si el cliente presenta créditos de consumo en los últimos 6 meses, mayor será la probabilidad de ser un cliente malo

Fuente: Hidalgo & Pichardo. Score de Bureau para cajas municipales de ahorro y créditos (2015) y Elaboración propia

- d. Impacto financiero en la implementación de un modelo Credit Scoring en una entidad de microfinanza

En la tabla V.4 los autores Chumacero & Sandoval (2011) presentan las siguientes variables:

**Tabla V.4 Impacto financiero en la implementación de un modelo Credit Scoring
en una entidad de microfinanzas**

Variable	Concepto
Edad	Edad del solicitante
Sexo y estado civil	Sexo y estado civil
Distrito	Distrito del titular
Antigüedad	Antigüedad del cliente
Máxima calificación 12M	Máxima calificación de los últimos 12 meses
Proporción normal 12M	Proporción de deuda en normal en los últimos 12 meses.
Número de entidades	Número de entidades con deuda
Incremento consecutivo saldo consumo 12M	Incremento consecutivo de saldo consumo en los últimos 12 meses
Decrecimiento consecutivo saldo consumo 12M	Decrecimiento consecutivo de saldo consumo en los últimos 12 meses.
Variación del número de entidades 12M	Variación del número de entidades en los últimos 12 meses
Variación del ratio Saldo Sistema excedente 12M	Variación del ratio saldo en el sistema financiero excedente en los últimos 12 meses.
Ratio saldo sistema / Máximo saldo sistema 12M	Ratio de saldo en el sistema financiero entre el máximo saldo en el sistema financiero en los últimos 12 meses
Máximo saldo en el banco respecto al sistema 6M	Máximo saldo en el banco respecto al sistema financiero en los últimos 6 meses.
Decrecimiento consecutivo Deuda Indirecta en el sistema 12M	Decrecimiento consecutivo deuda indirecto en el sistema financiero en los últimos 12 meses.

Fuente: Chumacero & Sandoval. Impacto financiero en la implementación de un modelo Credit Scoring en una entidad de microfinanzas (2011) y Elaboración propia

- e. Tesis doctoral A measurement of the small business Credit gap and the use of Credit Scoring by small financial

En la tabla V.5 el autor Hou (2016) presenta las siguientes variables:

Tabla V.5 A Measurement of the Small Business Credit Gap And The Use of Credit Scoring By Small Financial

Variable	Concepto
Written off	Variable ficticia, ya sea que el prestatario reembolse el préstamo o no
Orig_prin	La cantidad original de dinero que se prestó
Month_due	Términos del préstamo
Curr-int	Nivel de interés de los préstamos
Income amount	Nivel de ingreso anual de los préstamos
Housing	Variables ficticias, 1 es decir, los préstamos para el programa de vivienda; 0 significa el programa de préstamos para empresas
Gend	Género de los prestatarios
White	Variables dummy, si los prestatarios son personas blancas
Black	Variables dummy, si los prestatarios son personas negras
Lex	Variable ficticia, si los prestatarios viven en Lexington o no
CreditScore	Puntaje de crédito proporcionado por la agencia de crédito

Fuente: Hou. A measurement of the small business Credit gap and the use of Credit Scoring by small financial (2016) y Elaboración propia

- f. Tesis Small Business Credit Scoring: Evidence from Japan

En la tabla V.6, los autores Ryo & Hideaki (2010) presentan una aplicación del Credit Scoring en Japón, se utilizan las siguientes variables:

Tabla V.6 Small Business Credit Scoring: Evidence from Japan

Variable	Concepto
Independientes	Utilidad ordinaria/ entre pasivos totales
	Pasivos/Activos totales
	Efectivo/Activos corrientes
	Ganancias retenidas /Pasivos corrientes
	Utilidad bruta/Gasto total por intereses
	Ventas
	Total inventarios/ventas
Dependiente	La variable dependiente representa el riesgo de- fault (el Credit Scoring) y sería 0 en caso de su- pervivencia de la Pyme y 1 en caso de default.

Fuente: Ryo & Hideaki. Small Business Credit Scoring: Evidence from Japan (2010) y Elaboración propia

- g. Credit Scoring y la disponibilidad de crédito para pequeñas empresas en áreas con ingresos bajos y moderados (Credit Scoring and the Availability of Small Business Credit in Low- and Moderate-Income Areas)

En este documento se estima que el Credit Scoring está asociado con un aumento de \$3,900 en préstamos a pequeñas empresas en las muestras seleccionadas para áreas de bajos y moderados ingresos (LMI) atendidos, y este efecto es más o menos equivalente al estimado para áreas de mayores ingresos. En la tabla V.7, los autores Scott et al., (2004), utilizaron las siguientes variables:

Tabla V.7 Credit Scoring and the Availability of Small Business Credit in Low- and Moderate-Income Areas

Variable	Concepto
SBL (i, j)	La variable dependiente en nuestro modelo empírico es el volumen total de dólares de los préstamos a pequeñas empresas por debajo de \$100,000 (SBL (i, j)) para cada organización bancaria i, que respondió a la encuesta telefónica para cada tramo censal j en los Estados Unidos.
LNASSETS	Logaritmo de los activos
SBLRATIO	Ratio de préstamos para pequeñas empresas entre activos bancarios nacionales totales
EQUITY	Ratio total capital entre total activo
RURAL	Toma el valor de uno si es una zona rural
LNSBR	Logaritmo natural de los ingresos totales
HISPANIC	Proporción de los individuos identificados como hispanos
ASIAN	Proporción de los individuos identificados como asiáticos
BLACK	Proporción de los individuos identificados como negros
TRACTBRANCH	Si la institución bancaria tiene sucursal bancaria ubicada en el tramo censal.
MKTBRANCH	Si la institución bancaria tiene sucursal ubicada en el mercado bancario geográfico relevante.
SCORETRACT	Toma el valor de uno si la organización bancaria usa Credit Scoring para pequeñas empresas y mantiene una presencia física en tramo censal.
SCOREMKT	Toma el valor de uno si la organización bancaria usa Credit Scoring para pequeñas empresas y tiene una oficina fuera del tramo censal, pero dentro del mercado bancario.

Fuente: Scott et al., Credit Scoring and the availability of small business Credit in low- and moderate-income areas (2004) y Elaboración propia

h. Modelos de Credit Scoring (caso de Argentina):

Se construye un Credit Scoring con la técnica probit, los datos son obtenidos de CENDEU. El modelo pronostica el comportamiento de pago de los deudores personas y Pymes. En la tabla V.8, se resume lo que Gutiérrez (2007) propone.

Tabla V.8 Modelos de Credit Scoring – Matías Alfredo Gutiérrez Girault

Variable	Concepto
Calificación inicial	La calificación asignada al deudor al comienzo de cada período, que puede ser 1 o 2 por los filtros impuestos en la muestra historia Crediticia: aproximada por la calificación que tenía el deudor en la entidad 180 días antes de cada período.
Peor calificación sistema	La peor calificación que, al comienzo de cada período, tiene el deudor en el sistema financiero (puede o no coincidir con la calificación en la entidad).
Tipo de entidad	Es una variable categórica que busca controlar por distintos criterios o políticas en la evaluación del riesgo, que puedan afectar el perfil de riesgo de las carteras de préstamos. La variable identifica si el banco es público, de capital extranjero, etc.
Tipo de deudor	Busca controlar por tipo de deudor, consumo o vivienda o asimilable.
Clean	Es una variable dummy que toma un valor 0 si la deuda está cubierta con garantías preferidas A o B, y 1 de lo contrario. Por tratarse de deudores minoristas, es de esperar que refleje la existencia de garantías preferidas B, como es el caso de las financiaciones cubiertas con garantías prendarias o hipotecarias.
Total de bancos	Es una variable categórica que indica con cuántas entidades el deudor tiene obligaciones (1 o 2, 3 o 4, 5 o 6, 7 u 8, más de 8).
Crecimiento PBI	La tasa de crecimiento del PBI, durante cada período;
Deuda banco	Total de financiamiento tomado de la entidad, incluyendo responsabilidades eventuales.
Deuda sistema	Deuda total en el sistema financiero.
Deuda sistema	El cociente entre garantías preferidas y la deuda con la entidad
Significancia	El cociente entre la deuda con la entidad y la deuda total en el sistema, busca capturar la importancia que el endeudamiento con la entidad tiene en relación a su endeudamiento con el total del sistema financiero.

Fuente: Gutiérrez. Modelos de Credit Scoring (2007) y Elaboración propia

- i. Una aplicación del Credit Scoring en instituciones microfinancieras para el caso peruano propuesto por Salvador Rayo Cantón y la SBS “Diseño de un sistema de Credit Scoring para entidades de microfinanzas”:

En la tabla V.9, los autores Rayo (2010) & SBS (2010) propone las siguientes variables:

**Tabla V.9 Credit Scoring en instituciones microfinancieras para el caso peruano
propuesto por la SBS**

Variable
Créditos concedidos con anterioridad
Créditos concedidos en el último año
Número total de cuotas pagadas en el historial Crediticio
Número de cuotas incurridas en morosidad
Número de días de la mayor mora del cliente
Destino del Microcrédito, variable dicotómica (0) Capital de trabajo (1) Activo fijo.
Tipo de Operación de préstamo variable dicotómica (0) con incidencias (refinanciado, ampliado, etc.) (1) Normal
Tiempo de atención dedicada al cliente en términos de asesoría de inversiones, expresado en días.
Tipo de moneda en que concede el crédito.

Fuente: Rayo & Sbs. Variables Credit Scoring en instituciones microfinancieras para el caso peruano propuesto por la SBS (2010) y Elaboración propia

- j. Modelo de Credit Scoring Pymes para la medición del riesgo de morosidad de pequeña y mediana empresa – Un caso de aplicación en entidad financiera

Esta monografía es un estudio técnico que trata sobre una aplicación del Credit Scoring en Pyme. En la tabla V.10, los autores (Esquivel et al., 2013), proponen las siguientes variables:

Tabla V.10 Variables Modelo de Credit Scoring Pymes para la medición del riesgo de morosidad de pequeña y mediana empresa

Variable	Concepto
Cualitativas	Experiencia en el negocio
	Cartera de clientes
	Cantidad de competidores a nivel nacional
	Situación actual del mercado nacional
	Inmueble propio
Cuantitativas	Nivel de endeudamiento
	Margen de utilidad operativa
	Crecimiento en ventas
	Cobertura de intereses
	Solvencia
Nivel de Calificación del Scoring Pymes	Aceptable, se vislumbra que este cliente tendrá un buen comportamiento de pago.
	Opcional, aceptable situación financiera, no obstante presenta indicadores que deben ser estrictamente monitoreados.
	No aceptable, no es conveniente otorgarle un crédito pues el riesgo por morosidad que asume el banco es muy elevado.

Fuente: Esquivel et al., Modelo de Credit Scoring Pymes para la medición del riesgo de morosidad de pequeña y mediana empresa (2013) y Elaboración propia

k. Modelo de riesgo de crédito para las Pyme evidencia del mercado EE.UU (Modelling Credit risk for SMEs: Evidence from the US Market)

En la tabla V.11 los autores Altman & Sabato (2007), presentan las siguientes variables:

Tabla V.11 Modelling Credit risk for SMEs: Evidence from the US Market

Variable	Categoría de relación contable
Deuda a corto plazo / Capital (Valor en libros)	Apalancamiento
Efectivo / Total activos	Liquidez
EBITDA / Total activos	Rentabilidad
Ganancias retenidas / Total activos	Cobertura
EBITDA / Gastos por intereses	Actividad

Fuente: Altman & Sabato. Modelling Credit risk for SMEs: Evidence from the US Market (2007) y
Elaboración propia

5.2. Selección de variables

Las aplicaciones del Credit Scoring presentan diversas metodologías o funciones de enlace para modelar el riesgo de crédito. Siendo el modelo logístico utilizado en la banca por su fácil interpretación, estimaciones robustas y su buen desempeño en la implementación. En la mayoría de las aplicaciones del presente capítulo se ha utilizado la regresión logística.

Los modelos de admisión de clientes se aplican con mayor precisión en la aprobación de créditos, se calculan con variables externas de la entidad financiera, por ejemplo, datos del Reporte Crediticio Consolidado (RCC), información de Centrales de Riesgo, RENIEC, SUNAT, etc. Son eficientes en la prospección de nuevos clientes.

Una desventaja de los modelos de admisión es que en la medida que van ingresando nuevos clientes en el portafolio de créditos es posible que el comportamiento de la población cambie y se haga necesario calibrar el modelo.

Los modelos de comportamiento o behavior se usan para la gestión del portafolio de clientes y se construyen con las variables internas de la institución financiera, se

aplican a los clientes que están al día en sus pagos y los clientes con retraso en sus pagos. La fortaleza de estos modelos es que en su construcción recogen el comportamiento de la cartera de clientes y sirven para realizar acciones estratégicas con nuestros clientes. Por ejemplo, se aplican para el incremento o decremento líneas de crédito, realizar disposiciones de efectivo, mejorar tasas, para realizar acciones comerciales upselling y crossselling. Apoyan a la gestión de cobranzas en las etapas de mora temprana hasta los 60 días.

Los Modelos de Buró, utilizan variables externas del Reporte Crediticio Consolidado (RCC), información de centrales de riesgo, información de impuestos (SUNAT), RENIEC, ingresos, archivo de datos negativos, información de pago de servicios, información de empresas de comunicaciones, etc. Lo que lo diferencia de otros modelos es la profundidad histórica que es mayor a 36 meses, muchas instituciones financieras y la SBS utilizan 5 años de historia. Se busca siempre estabilidad en las variables del modelo y se mide el performance de un cliente en una ventana de predicción de 6 meses o un año. Previamente a su construcción se hace una diferenciación por tipo de crédito y por tipo de persona natural o jurídica ya que los niveles de riesgo son diferentes. La desventaja de estos modelos es que demandan de mucha información en línea y su tiempo de procesamiento es relativamente alto. Se aplican en todas las etapas de un crédito y generalmente llegan a cubrir con un score a los 9 millones de clientes bancarizados en la RCC.

Los modelos de cobranza son intensivos en información diaria ya que día a día caen en mora nuevos clientes, y también diariamente los clientes pagan sus obligaciones por lo que son modelos dinámicos, sensibles a la variable de pagos, generalmente se construyen por tramos de mora, desde el tramo cero o al día, tramo 1- 30 días de atraso, hasta los estadios de mora tardía, cobranza judicial y créditos castigados. En cada etapa de morosidad los clientes presentan diferentes comportamientos y se utilizan diferentes estrategias de recuperación que se apoyan en los modelos de cobranzas. Los modelos de recuperación de castigos se utilizan en los estadios de mayor morosidad y se construyen con data interna de la institución financiera. Los modelos de cobranza se aplican en la valorización de cartera, donde nos ayudan a establecer la probabilidad de recuperación en cada tramo de mora. Dichos modelos sirven para el cálculo de la rentabilidad

de la empresa, en la definición del capital regulatorio de la institución financiera, en la construcción del CLV (Customer Lifetime Value) y PLV (Prospect Lifetime Value) de la compañía, entre otras muchas aplicaciones, en nuestro país están tomando relevancia ya que su uso en las empresas está generando beneficios.

Los modelos de Rating generalmente se utilizan en la banca corporativa y en los segmentos de alto endeudamiento y alto nivel de ventas, donde predomina la formalidad de las empresas, el buen gobierno corporativo y la transparencia de la información, ya que utilizamos los indicadores financieros para construir los modelos de Rating. Una de las desventajas es que se cuentan con poco default de clientes, ya que estamos en un segmento estable. Estos modelos no pueden ser aplicados en las Pymes por su nivel de informalidad.

El Credit Scoring que se va a desarrollar se diferencia de las aplicaciones revisadas en los siguientes aspectos:

- ✓ Es un modelo de admisión de clientes, es decir es muy preciso en la aprobación de créditos, el modelo se aplica solo a personas que tienen una experiencia previa en el sistema financiero.
- ✓ Se recoge información del dueño de la Pyme y de los intervinientes y se aplica en el producto capital de trabajo en la modalidad revolvente, porque es el producto más solicitado. Las aplicaciones revisadas no diferencian entre clientes revolvente y no revolvente, son más genéricos.
- ✓ Es construido con información interna de la institución financiera, por lo tanto, no es un modelo genérico tipo buró, en su construcción no se utilizan otros modelos como imput o variable independiente, cosa que si sucede con algunos de los modelos revisados en las aplicaciones.
- ✓ Para construir el modelo la profundidad histórica de las variables es de 24 meses, detalle que no se observa en ninguno de los modelos revisados. Siendo la ventana de predicción de un año lo que implica estimaciones estables para la entidad financiera.

- ✓ Las variables que son comunes en el análisis son por ejemplo la zona geográfica, la clasificación en el sistema financiero, el número de entidades donde el cliente tiene deuda, la deuda en el sistema financiero, los días de mora en sistema financiero y la mora en la institución financiera, nivel de endeudamiento.
- ✓ En los modelos revisados se observa que no se explota la información del titular y los socios de la Pyme, en nuestro caso si consideramos dicha información y además diferenciamos a las personas naturales y jurídicas.

En el Perú, las Pyme tienen problemas de informalidad, debido a los altos costos de formalización y el desconocimiento de los beneficios que tiene una empresa formal. Cuando se evalúa una Pyme, el funcionario de negocio en la visita de campo genera un balance, un inventario, mide el stock de acuerdo a lo que observa, evalúa las garantías y los ingresos del negocio. El problema radica en que en la mayoría de casos esta información es poco confiable ya que la data generada es aproximada, es por ello que muchas veces la información financiera de la Pyme es descartada, sin embargo, la institución financiera viene implementando controles de calidad a sus evaluaciones Crediticias, validando dicha información mediante auditorias periódicas de su proceso de admisión. Es por ello que utilizaremos algunas variables con información financiera de la Pyme.

Una ventaja del sistema financiero peruano es el Reporte Crediticio de Consolidado (RCC), que contiene la información de las deudas a nivel documento nacional de identidad y cuenta contable (información de todos los productos en todas instituciones financieras) de todas las personas que han solicitado un crédito. Dicha base de datos es distribuida por la SBS a todas las instituciones financieras y centrales de riesgo, es por ello que utilizaremos variables de la RCC para la construcción del Credit Scoring.

Los clientes bancarizados y que sean nuevos en la institución financiera, dado que no tenemos información interna se usa la RCC.

Los clientes no bancarizados no son el objetivo de la institución financiera, por eso no son atendidos masivamente mediante un Credit Scoring, en algunos casos se otorga un crédito por excepción y solicitándole al cliente garantías reales. Todos los casos de

excepción están contemplados en la política de créditos y materializan mediante reglas de negocio. Posteriormente, a los clientes admitidos por excepción se les aplica el Credit Scoring en la etapa de seguimiento del crédito.

De acuerdo a la disponibilidad de la información del sistema financiero y la institución financiera se tienen las siguientes variables en la tabla V.12, donde se aprecia los siguientes grupos de variables: Información del solicitante, experiencia previa en la institución financiera, información de los productos activos y pasivos en la institución financiera, experiencia en el sistema financiero, información de vinculación con la institución financiera y la información financiera del solicitante.

Tabla V.12 Construcción de Variables

Grupos de variables	Variable
Información del solicitante	Edad
	Sexo
	Ubicación geográfica
	Meses de constitución
	Intervinientes
	Persona Natural jurídica
	Garantías reales
Experiencia previa en la institución financiera	Atraso de los intervinientes en los últimos 12 meses
	Máximo atraso del titular en los últimos 3 meses
Información de productos de activo y pasivo	Ratio entre saldo de activo del último mes y el máximo saldo de activo en los últimos 24 meses de todos los intervinientes en la institución financiera
	Saldo promedio de pasivo en los últimos 12 meses de todos los intervinientes en la institución financiera
	Saldo promedio de pasivo en los últimos 6 meses de todos los intervinientes
	Saldo de Pasivo mínimo en los últimos 3 meses de todos los intervinientes en la institución financiera
	Saldo pasivo mínimo en los últimos 6 meses del titular en la institución financiera
	Ratio entre saldo de activo del último mes y el promedio de saldo de activo en los últimos 12 meses del titular

Experiencia en el sistema financiero	Ratio entre saldo de activo últimos 3 meses y saldo de activo últimos 24 meses de todos los intervinientes de la institución financiera
	Saldo de activo vencido máximo en los últimos 12 meses del titular en la institución financiera
	Cantidad de empresas del SF en que ha sido reportada el titular
	Diferencia entre meses con buen comportamiento y meses con mal comportamiento en el SF en los últimos 24 meses de todos los intervinientes
	Diferencia entre meses con buen comportamiento y meses con mal comportamiento en el SF en los últimos 6 meses de todos los intervinientes
	Número de meses con sobregiro en los últimos 6 meses en el sistema financiero del titular
	Número de meses desde que algún interviniente tuvo clasificación CPP o peor en el último año
	Saldo de endeudamiento último mes/saldo endeudamiento promedio en los últimos 12 meses de todos los intervinientes
	Saldo de endeudamiento últimos 3 meses/ saldo endeudamiento promedio en los últimos 24 meses del titular
	Saldo de endeudamiento últimos 3 meses/ saldo endeudamiento promedio en los últimos 24 meses de todos los intervinientes
	Saldo de endeudamiento último 6 meses / saldo endeudamiento promedio en los últimos 24 meses del titular
	Peor clasificación de riesgo del titular en los últimos 12 meses
	Peor clasificación de riesgo de todos los intervinientes en los últimos 24 meses
Información de vinculación con la institución financiera	Antigüedad Máxima del activo o pasivo de los intervinientes en la institución financiera
	Antigüedad máxima del activo o pasivo vigente de los intervinientes en la institución financiera
Información financiera del solicitante	Capital de trabajo neto
	Monto Total Patrimonio
	Ratio entre pasivo y Patrimonio
Fuente: Capítulo V y Elaboración propia	

CAPÍTULO VI. APLICACIÓN DEL CREDIT SCORING EN PYMES

La relación que existe entre capital de trabajo y el éxito de las Pymes, es directamente proporcional, sin el capital de trabajo necesario la pequeña y mediana empresa no puede atender a sus clientes ni cubrir sus necesidades operativas, con lo cual no generarían beneficios.

El producto capital de trabajo tiene dos modalidades revolvente y no revolvente, la institución financiera tiene como estrategia ofertar capital de trabajo en la modalidad revolvente, ya que es uno de los productos más demandados, es por ello que se propone construir un modelo de Credit Scoring aplicándolo a la admisión de clientes del producto capital de trabajo en la modalidad revolvente.

6.1. Aplicación del Credit Scoring en el producto Capital de Trabajo Revolvente

Actualmente las entidades financieras vienen implementando el nuevo acuerdo de capital Basilea II y III, en este contexto la medición, el control y la gestión de los riesgos se orientan al desarrollo de modelos internos Credit Scoring (Bis 2019).

La institución financiera necesita mejorar la gestión del riesgo de crédito y el cálculo de los requerimientos de capital regulatorio y provisiones por riesgo Crediticio.

Es por ello que para conocer el riesgo de los clientes que solicitan tener un producto capital de trabajo en la modalidad revolvente se necesita contar con un sistema de calificación adecuado que permita distinguir a los clientes más riesgosos de los menos riesgosos.

El Credit Scoring para Pyme revolvente, es una herramienta estadística que se construye a partir de información histórica; identifica patrones y el perfil de riesgo de clientes Pyme a través de variables demográficas, información de la visita de campo, comportamiento en el sistema financiero y operativa previa de activo y pasivo.

El Credit Scoring propuesto es construido empleando métodos estadísticos de regresión logística.

De esta forma se le asigna una probabilidad a la operación o solicitud de los clientes Pyme capital de trabajo en la modalidad revolvente, dicho score servirá para aprobar o denegar las solicitudes generadas.

El modelo debe implementarse en una aplicación que permita ejecutarlo durante el proceso de evaluación de solicitudes de crédito, la implementación en un sistema de la institución financiera no es parte del alcance de la presente tesis.

Es importante para fines de seguimiento, validación y aplicación en la gestión, es necesario que las variables finales del modelo deban ser almacenadas en una herramienta de base de datos, según las recomendaciones de Basilea Bis (2010).

Para la construcción de un modelo de Credit Scoring se realizan los siguientes pasos:

6.1.1 Definición de la Población Objetivo

La población objetivo son las Pymes que soliciten financiamiento para capital de trabajo en la modalidad revolviente.

El modelo no aplica en los siguientes casos:

- ✓ Cliente No Pyme
- ✓ Solicitud de activo fijo Pyme
- ✓ Solicitud de campañas permanentes o estacionales
- ✓ Solicitud de operación leasing Pyme

6.1.2 Limitaciones del Modelo

Para la correcta aplicación del modelo solo se debe utilizar para solicitudes de productos Pyme revolviente para capital de trabajo. La institución financiera no trabaja con clientes sin experiencia en el sistema financiero, salvo excepciones en su política de créditos.

6.1.3 Variables del Credit Scoring revolving:

La tabla VI.1, se aprecia algunas de las variables significativas para el Credit Scoring revolving, dichas variables pasaron los análisis univaridos, correlaciones y bivariados.

Tabla VI.1 Variables Credit Scoring Revolvente

Tipo de Información	Variable
Endeudamiento en la Institución Financiera	Ratio entre saldo de activo últimos 3 meses y saldo de activo últimos 24 meses de todos los intervinientes.
Endeudamiento sistema financiero	Cantidad de empresas del Sistema Financiero en que ha sido reportada el titular. Meses con sobregiro en los últimos 6 meses en el sistema financiero del titular Evolución de endeudamiento en el Sistema Financiero de los últimos 3 reportes vs los últimos 24 reportes de todos los intervinientes con ventana fija.
Información financiera	Ratio entre pasivo y patrimonio Capital de trabajo neto Monto total patrimonio
Mora en la Institución Financiera	Saldo de activo vencido máximo en los últimos 12 meses del titular
Mora sistema financiero	Diferencia entre meses con buen comportamiento y meses con mal comportamiento en los últimos 24 reportes de la RCC de todos los intervinientes. Peor clasificación de riesgo de todos los intervinientes en los últimos 24 reportes de la RCC.
Pasivo en la Institución Financiera	Saldo de pasivo mínimo en los últimos 6 meses del titular.
Solicitante	Zona Meses transcurridos desde la constitución de la Pyme

Fuente: Capítulo VI y Elaboración propia

6.1.4 Pruebas de coherencia

Una vez que tenemos las variables del modelo y su forma de cálculo, es necesario mencionar que valores pueden tomar cada variable para tener consistencia con los valores que toma. En la tabla VI.2, se aprecia 10 ejemplos de pruebas de coherencia.

Tabla VI.2 Pruebas de coherencia

Nº	Descripción de la variable	Coherencia
1	Número de meses en que todos los intervinientes registraron atrasos en la institución financiera en los últimos 3 meses	Se espera valores entre 0 y 12
2	Peor atraso en días del titular en la institución financiera en los últimos 3 meses.	No se espera valores negativos
3	Cantidad de empresas del SF que reportaron al titular en el mes previo a la solicitud	No se espera valores negativos
4	Diferencia entre el número de meses con buen y mal comportamiento de todos los intervinientes en el SF durante los últimos 24 meses	Se espera valores entre -24 y 24
5	Diferencia entre el número de meses con buen y mal comportamiento de todos los intervinientes en el SF durante los últimos 6 meses	Se espera valores entre -6 y 6
6	Numero de meses con sobregiro del titular en los últimos 6 meses	Se espera valores entre 0 y 6
7	Número de meses transcurridos desde el último reporte RCC en que alguno de los intervinientes tuvo clasificación CPP o peor en el último año.	Se espera valores entre 0 y 13
8	Ratio de deuda promedio SF de todos los intervinientes del último mes respecto a los últimos 12 meses considerando ventanas fijas para el cálculo.	No se espera valores negativos
9	Ratio de deuda promedio SF del titular de los últimos 3 meses respecto a los últimos 24 meses considerando ventanas fijas para el cálculo.	No se espera valores negativos
10	Ratio de deuda promedio SF de todos los intervinientes de los últimos 3 meses respecto a los últimos 24 meses considerando ventanas fijas para el cálculo.	No se espera valores negativos

Fuente: Capítulo VI y Elaboración propia

6.1.5 Definición de Incumplimiento

En la tesis nos referiremos a default o incumplimiento cuando la operación haya registrado un atraso mayor a 60 días en los 12 meses posteriores a la concesión del crédito.

Según la resolución SBS N° 11356 si un cliente pasa los 60 días de atraso, la provisión que realiza la entidad financiera es el 60% del saldo deudor, mientras que si el atraso esta entre 31 y 60 días la provisión es del 25% SBS (2008).

Finalmente, los gastos de provisión que se generan por los incumplimientos impactan directamente a la utilidad en el estado de ganancias y pérdidas de la entidad financiera. En la figura VI.1, se aprecia como un cliente cae en incumplimiento en el mes séptimo, en dicho momento presenta más de 60 días de atraso. La medición se realiza en una ventana de observación de 12 meses posteriores a la concesión del crédito.

Figura VI.1 Default vista 12 meses



Fuente: Capítulo VI y Elaboración propia

Default= 1 Si la Pyme se atrasa más de 60 días en los siguientes 12 meses.

Default= 0 Si la Pyme no se atrasa más de 60 días en los siguientes 12 meses

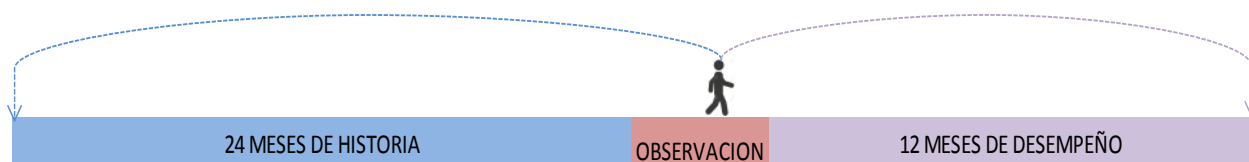
6.1.6 Ventana de Observación

Por motivos de confidencialidad de la información en la tesis trabajaremos con información de una muestra histórica de datos aleatorios tomados de la institución financiera con una cierta antigüedad de años. Al ser una muestra aleatoria las inferencias que se van a estimar son estadísticamente significativas.

- ✓ La ventana de observación esta entre el 1 de julio de 2005 y el 30 de abril de 2009.
- ✓ La profundidad histórica de las variables es de 24 meses. Quiere decir que para cada mes de observación vamos a tomar información con una profundidad de 24 meses.
- ✓ La ventana de predicción o desempeño es de 12 meses. Es decir, vamos a ver el comportamiento de la Pyme hasta el 30 de abril de 2010.

En la figura VI.2, se muestra a un cliente al cual evaluamos con una profundidad histórica de 24 meses, su ventana de observación es el mes evaluación del crédito y su ventana de desempeño es de 12 meses. Dicho esquema se repite mensualmente para todas las unidades de análisis en el modelo.

Figura VI.2 Ventana de Observación del Credit Scoring



Fuente: Capítulo VI y Elaboración propia

Si se quiere calibrar el Credit Scoring, se debe utilizar la definición regulatoria de la resolución N°14354 SBS (2009), que establece que para exposiciones minoristas, la probabilidad de default debe calcularse a nivel operación.

6.1.7 Segmentación

Se utiliza como criterio de segmentación el tipo de producto solicitado, en este caso tomaremos el producto capital de trabajo en la modalidad revolvente. Además, la estrategia de la institución esta direccionada en crecer en líneas revolventes, que según el comportamiento de pagos son menos riesgosas y son cada vez más solicitadas en el mercado.

En la bibliografía revisada a nivel de productos Pyme que ofrecen las instituciones financieras del Perú se observó que uno de los principales productos es el capital de trabajo en la modalidad revolvente.

6.1.8 Características de la muestra

En el presente trabajo se tiene información recogida de una muestra aleatoria de los años 2005 hasta el 2010.

En la tabla VI.3, se aprecia que la muestra de entrenamiento (train) es de 15,904 operaciones revolventes, de las cuales 7,404 son persona jurídica y 8,497 son persona natural. La muestra de test es de 5,303 operaciones revolventes, de las cuales 2,486 son persona jurídica y 2,817 son persona natural.

Tabla VI.3 Muestras de operaciones revolventes

Muestra	Jurídico	Natural	Total
Test	2,486	2,817	5,303
Train	7,407	8,497	15,904
Total	9,893	11,314	21,207

Fuente: Capítulo VI y Elaboración propia

En la tabla VI.4, se aprecia el detalle las distribuciones de las operaciones morosas:

Tabla VI.4 Proporción de operaciones morosas por muestra

Distribución de las Muestras						
Modelo	Construcción	Nº Morosas	% Morosas	Test	Nº Morosas	% Morosas
Revolvente	15,904	1,898	11.93%	5,303	633	11.94%

Fuente: Capítulo VI y Elaboración propia

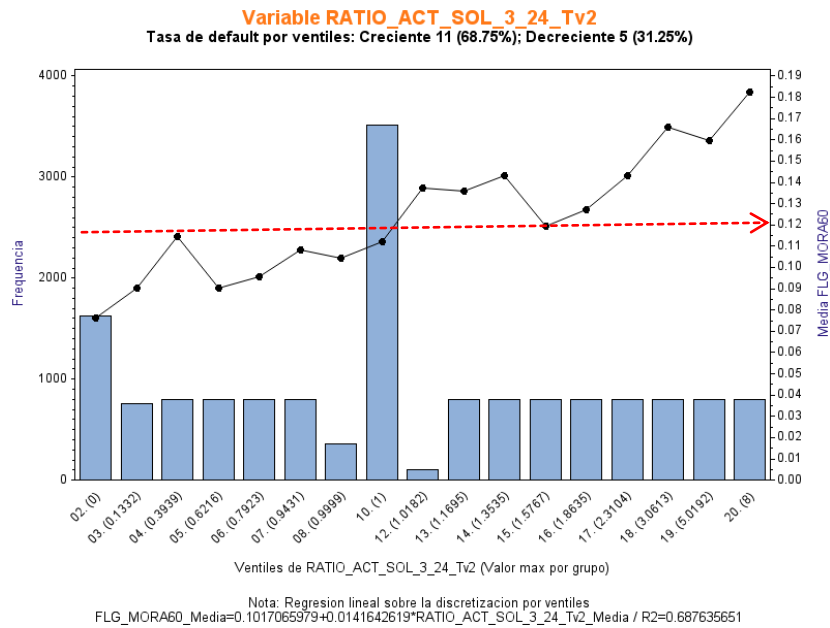
6.2. Principales características que influyen en el modelo Credit Scoring

Se analiza más de 500 variables candidatas a ingresar al modelo, se han definido grupos o tipos de variables, para evitar la correlación entre los mismos tipos de variables, a la vez dentro de cada grupo o tipo buscamos a la variable que tenga mayor correlación con el default. Tenemos grupos o tipos de variables que explican endeudamiento, endeudamiento en el sistema financiero, información financiera, mora, mora en el sistema financiero, pasivos, información del solicitante y vinculación con la entidad. El proceso de análisis descriptivo, univariado, bivariado se inicia con más de 500 variables en diferentes ventanas de tiempo, pero van descartándose variables por problemas de concentración, porcentaje elevado de valores vacíos, por elevada correlación entre variables de un mismo tipo y por no presentar una relación directa o indirecta con el default, esto se traduce en una contradicción en los signos esperados de cada variable. Al pasar los análisis iniciales se realiza el tratamiento de datos y las pruebas de coherencia y nos quedamos con las variables que tienen más sentido de negocio y las más predictivas. En cada etapa de la modelación se toma en cuenta la relación creciente o decreciente de cada una de las variables con el incumplimiento de pago mayor a 60 días.

En el caso de la variable garantías no todos los clientes de la institución tenían garantías reales es decir un deposito a plazos o un bien inmueble o mueble, solo se tenían en garantía herramientas de trabajo y eso no constituye una garantía real, es por ello que la variable no entra al modelo. Para las variables edad y sexo se observó que dicha información no era registrada correctamente en la solicitud de crédito en muchos casos se obviaba y esto generó un alto porcentaje de valores vacíos es por ello que dichas variables no ingresan al modelo. Sin embargo, dichas variables pueden ingresar en una etapa de calibración del modelo (etapa del ciclo de vida de modelo donde se hacen ajustes al modelo para un perfil de cliente específico, por ejemplo, Pymes nuevas) o en el caso de los clientes con garantías reales se aplican mejores condiciones Crediticias, incremento de líneas, croselling, consolidación de deuda, etc.

Se muestra el análisis de las variables relevantes en los diferentes grupos o tipos de información analizados.

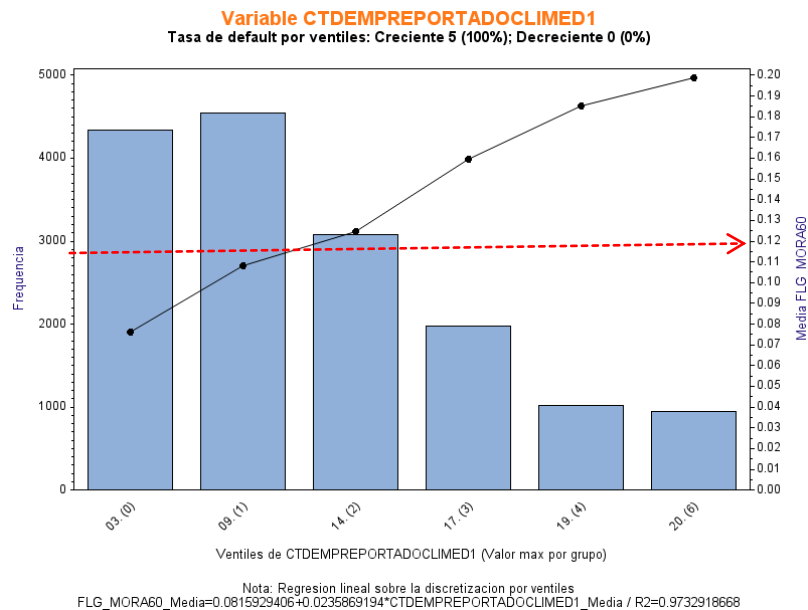
Figura VI.3 Tasa de Default por ventiles ratio de endeudamiento



Fuente: Capítulo VI y Elaboración propia

En la figura VI.3, se aprecia que, si aumenta el saldo activo de los últimos tres meses entre saldo de activo de los últimos 24 meses entonces el riesgo de default aumenta.

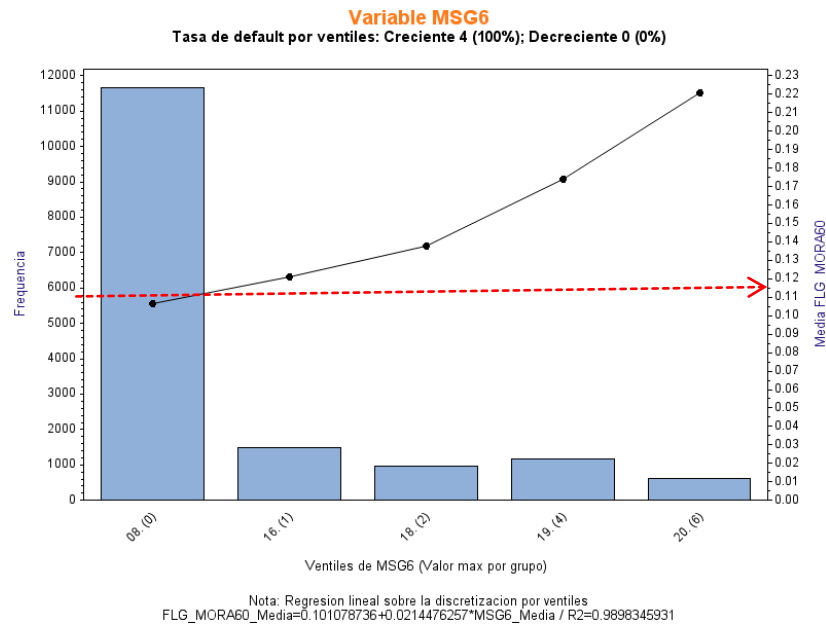
Figura VI.4 Tasa de default por ventiles de número de empresas reportado



Fuente: Capítulo VI y Elaboración propia

En la figura VI.4, se aprecia que, si aumenta el número de empresas que reportan al titular de la Pyme entonces aumenta el riesgo de incumplimiento.

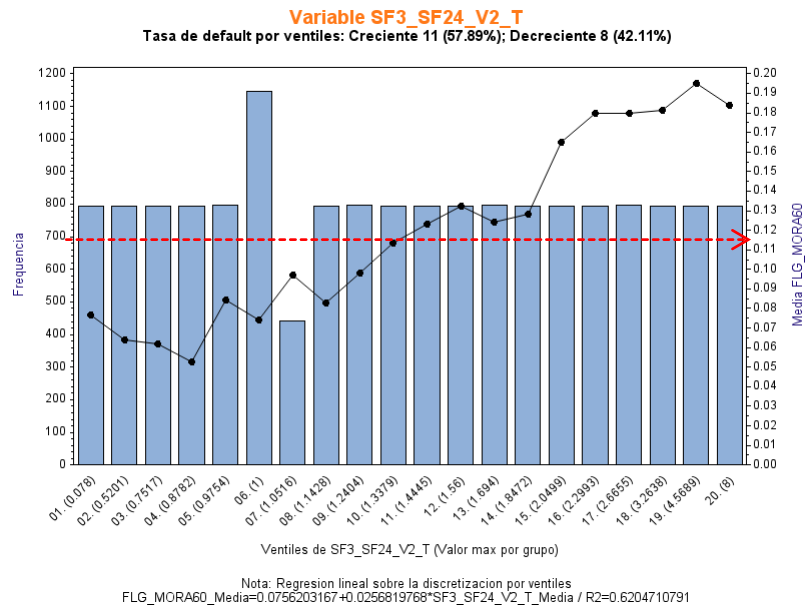
Figura VI.5 Tasa de default por ventiles de sobregiro



Fuente: Capítulo VI y Elaboración propia

En la figura VI.5, se aprecia que, si una Pyme incrementa la cantidad de meses con sobregiro entonces la tasa de incumplimiento aumenta.

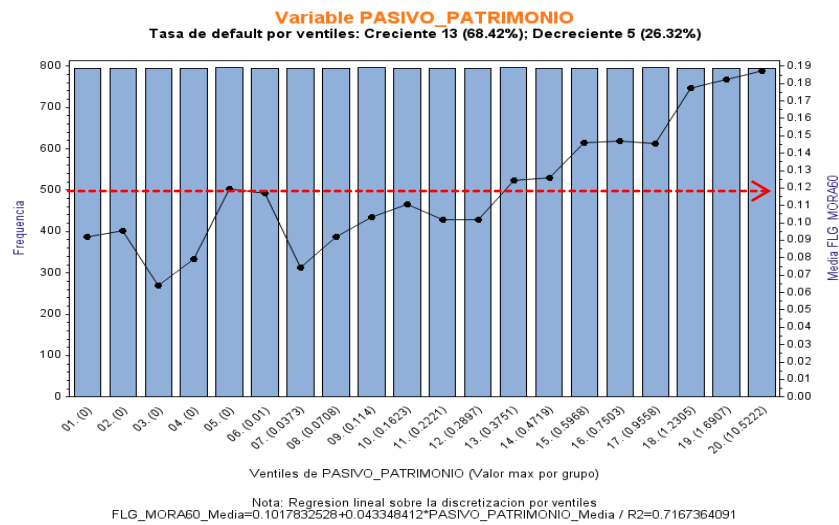
Figura VI.6 Tasa de default por ventiles evolución de deuda en el sistema



Fuente: Capítulo VI y Elaboración propia

En la figura VI.6, se aprecia que, si la Pyme aumenta el nivel de endeudamiento entonces el riesgo de incumplimiento crece.

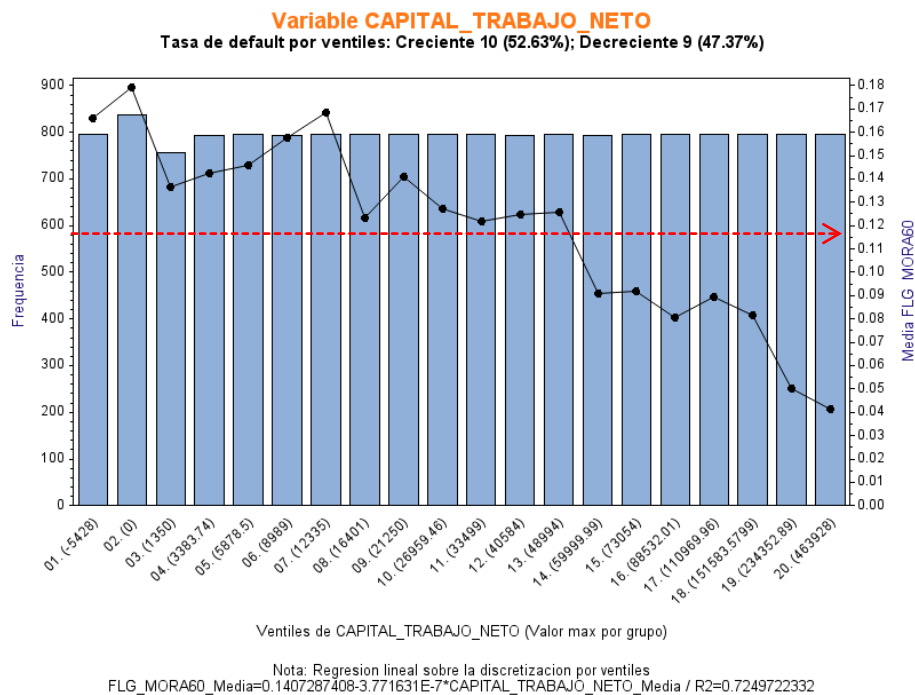
Figura VI.7 Tasa de default por ventiles pasivo entre patrimonio



Fuente: Capítulo VI y Elaboración propia

En la figura VI.7, se aprecia que, si la Pyme toma mayor apalancamiento el riesgo de incumplimiento aumenta. La tasa de incumplimiento promedio igual a 11.9%.

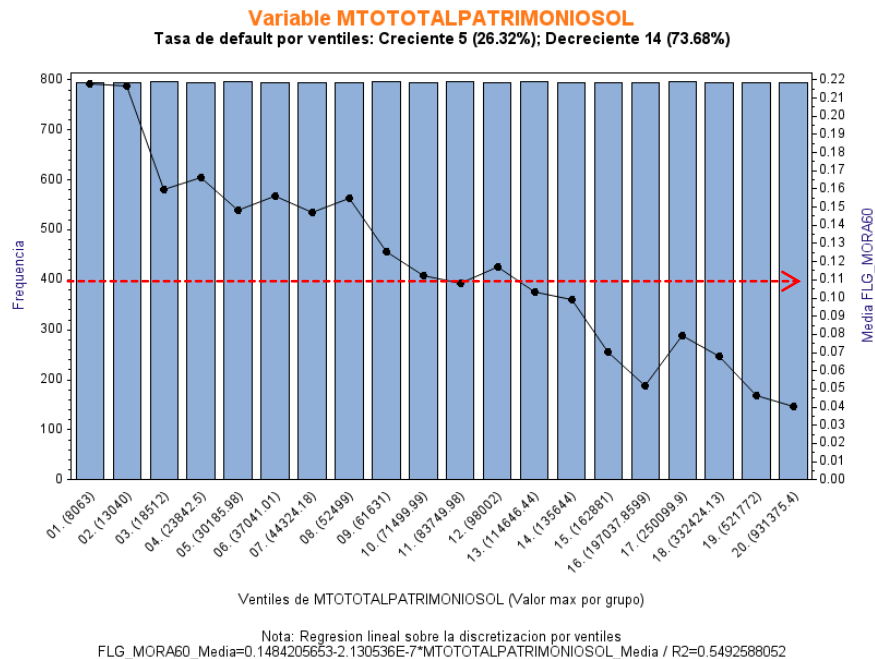
Figura VI.8 Tasa de default por ventiles capital de trabajo neto



Fuente: Capítulo VI y Elaboración propia

En la figura VI.8, se aprecia que, si el capital de trabajo neto, es decir la liquidez de la Pyme se incrementa entonces la tasa de incumplimiento disminuye.

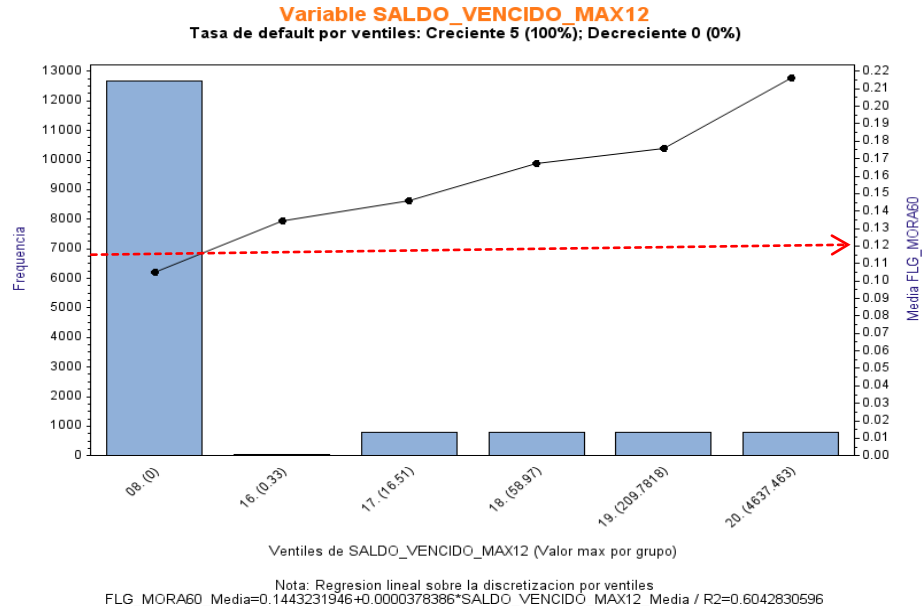
Figura VI.9 Tasa de default por ventiles de patrimonio



Fuente: Capítulo VI y Elaboración propia

En la figura VI.9, se aprecia que, si aumenta el patrimonio el riesgo disminuye.

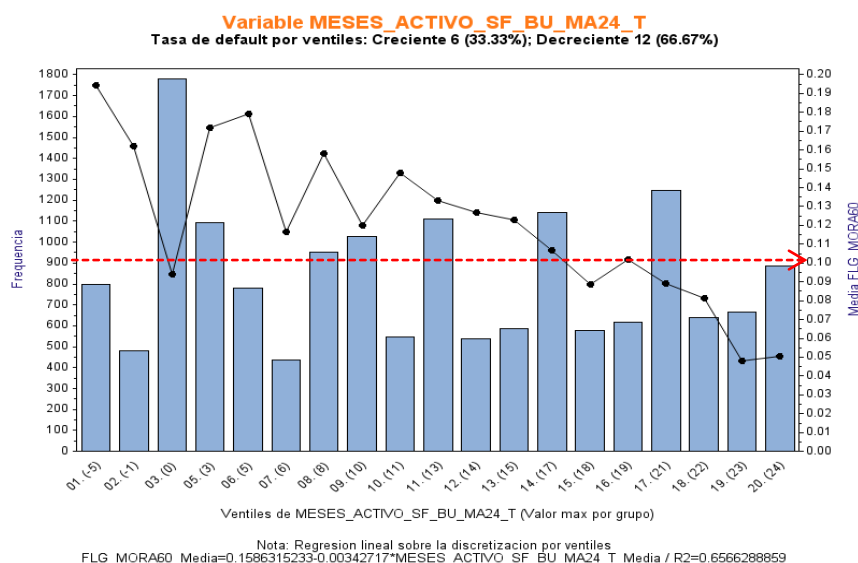
Figura VI.10 Tasa de default por ventiles saldo activo vencido



Fuente: Capítulo VI y Elaboración propia

En la figura VI.10, se aprecia que, si aumenta el saldo activo vencido entonces el riesgo aumenta.

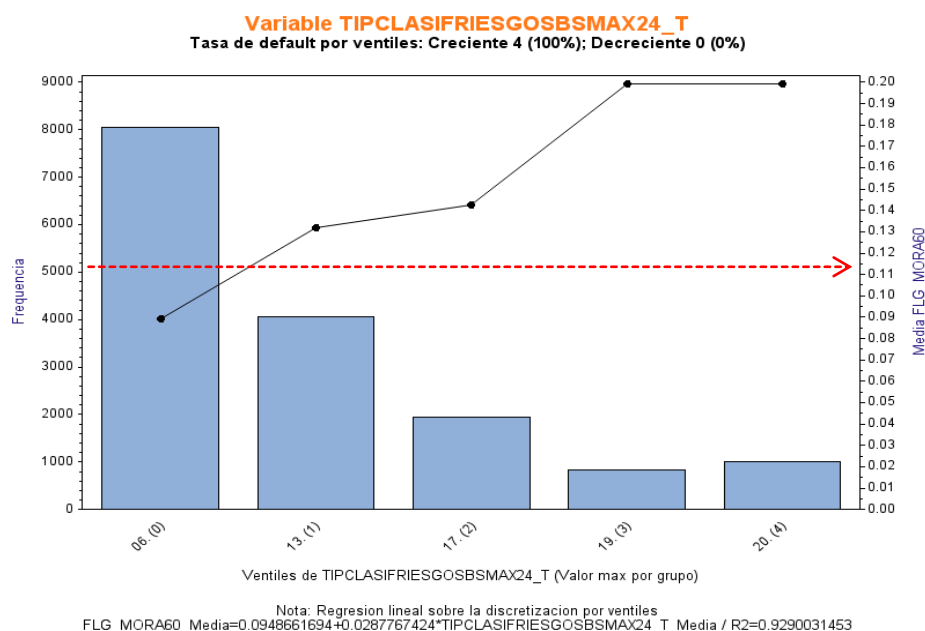
Figura VI.11 Tasa de default por ventiles de diferencia entre meses con buen comportamiento y mal comportamiento en el sistema financiero



Fuente: Capítulo VI y Elaboración propia

En la figura VI.11, se aprecia que, si la diferencia entre meses con buen y mal comportamiento se incrementa entonces la tasa de incumplimiento disminuye.

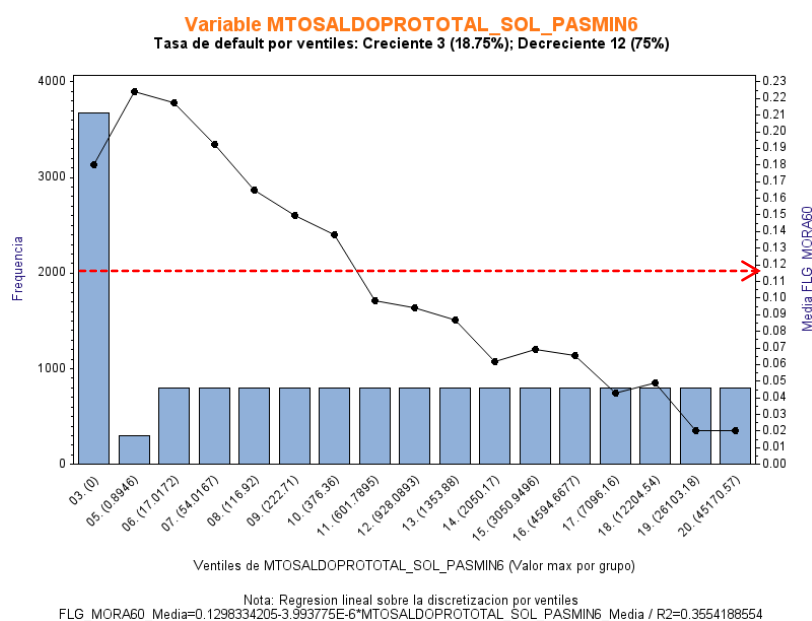
Figura VI.12 Tasa de default por ventiles peor clasificación de riesgo en los últimos 24 reportes en el sistema financiero



Fuente: Capítulo VI y Elaboración propia

En la figura VI.12, se aprecia que, si la Pyme tiene peor clasificación en sistema financiero entonces la tasa de incumplimiento aumenta.

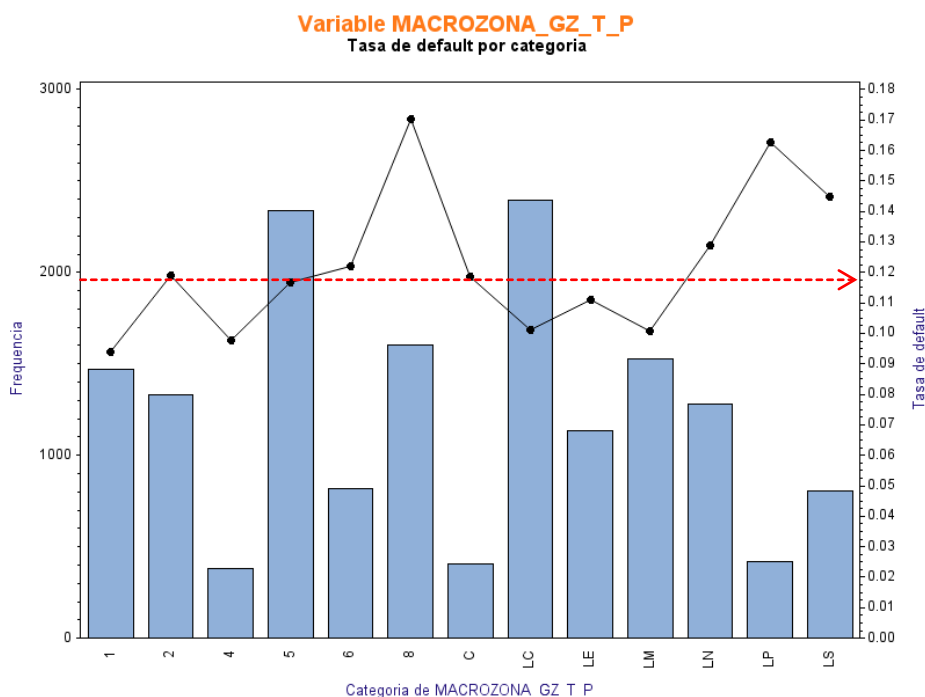
Figura VI.13 Tasa de default por ventiles del pasivo



Fuente: Capítulo VI y Elaboración propia

En la figura VI.13, se aprecia que, si aumentan los pasivos de la Pyme entonces la tasa de incumplimiento disminuye.

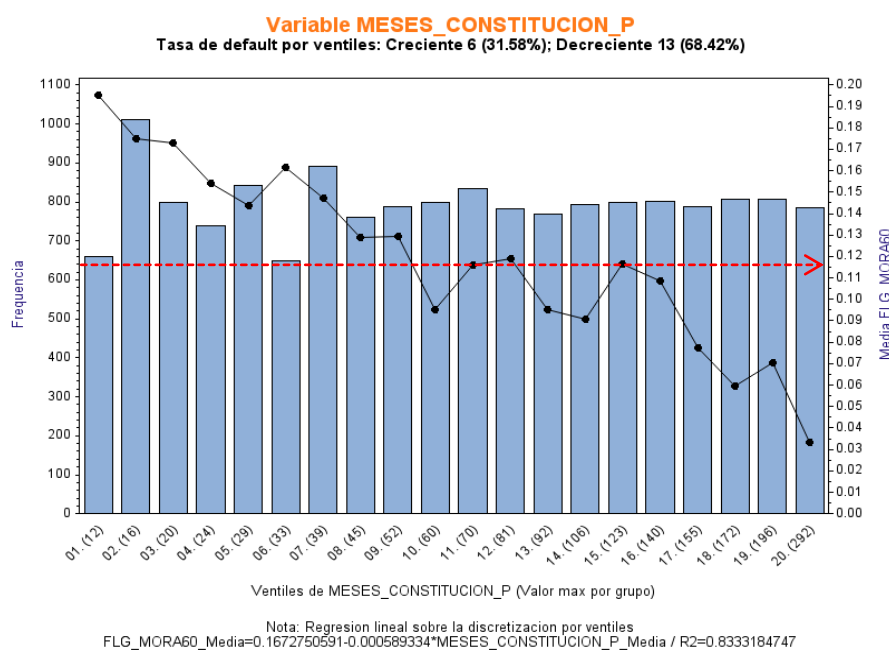
Figura VI.14 Variable zona geográfica



Fuente: Capítulo VI y Elaboración propia

En la figura VI.13, se aprecia que, si las zonas son de mayor riesgo (como por ejemplo Callao) la tasa de incumplimiento aumenta.

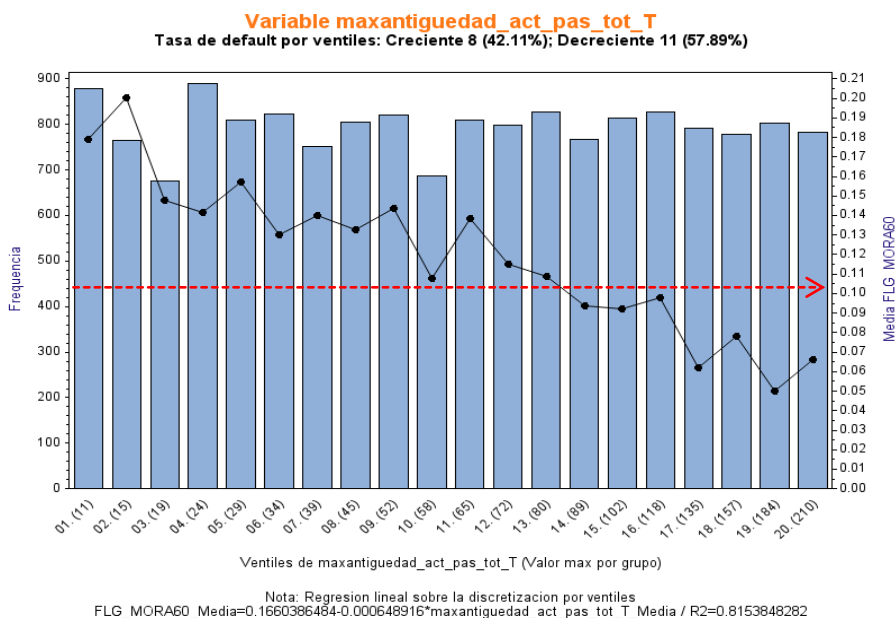
Figura VI.15 Tasa de default por ventiles de meses de constitución



Fuente: Capítulo VI y Elaboración propia

En la figura VI.15, se aprecia que, si la Pyme tiene más meses de constitución en el mercado entonces el riesgo de incumplimiento disminuye.

Figura VI.16 Tasa de default por ventiles de antigüedad



Fuente: Capítulo VI y Elaboración propia

En la figura VI.16, se aprecia que, si la Pyme tiene más antigüedad en la institución financiera entonces el riesgo de incumplimiento disminuye.

6.3. Credit Scoring Revolvente

Se aplica la función de enlace logit o regresión logística, según Guevara et al., (2017) se usan los modelos logit en el mayor porcentaje de instituciones financieras.

Las variables que ingresan al Credit Scoring deben cumplir con dos condiciones básicas para ser seleccionadas: ser significativas dentro de la regresión, lo cual implica que aporten capacidad discriminante y no sean redundantes y otra condición es que ninguna variable concentre un peso excesivo. El peso en el modelo se calcula a través del estadístico de Wald (valor del estadístico de la variable/suma de valores del estadístico de todas las variables incluidas en el modelo final).

El proceso de obtención del modelo final es:

1. Para construir el modelo se utilizó el 75% de la muestra para entrenar (intramuestral - train) y 25% para el test (fuera de muestra).
2. Se construyeron regresiones con combinaciones de todas las variables que pasaron a la etapa de multivariado, es decir los análisis univariados, bivariados, signos esperados, imputación de valores faltantes, correlaciones, etc.
3. Se utilizó el proceso de selección estadística llamado Stepwise.

El método de selección Stepwise permite ingresar las variables iterativamente a través de criterios de significación como la prueba F, finalizando cuando alguna variable candidata deja de cumplir con la condición de ingreso. En este método se consideró un nivel de confianza del 95% ($p \text{ valor} \leq 0.05$).

Finalmente, se seleccionó el modelo con mayor poder predictivo (Gini), cuyas variables cumplen las condiciones estadísticas adecuadas y tienen sentido de negocio.

En la tabla VI.5, se presenta el Credit Scoring revolvente para capital de trabajo:

Tabla VI.5 Modelo Revolvente

Modelo Revolvente		ROC=75.7%, Gini=51.5%	
Grupo	Variable	Peso	Peso Grupo
Endeudamiento en la institución financiera	Ratio entre saldo de activo últimos 3 meses y saldo de activo últimos 24 meses de todos los intervinientes con ventana fija	1.6%	1.6%
	Endeudamiento sistema financiero	19.6%	35.6%
	Meses con sobregiro en los últimos 6 meses en el sistema financiero del titular	6.6%	
	Evolución de endeudamiento en el Sistema Financiero de los últimos 3 reportes vs los últimos 24 reportes de todos los intervinientes con ventana fija.	9.5%	
Información financiera	Ratio entre pasivo y patrimonio	1.1%	10.0%
	Capital de trabajo neto	2.3%	
	Monto total patrimonio	6.6%	
Mora en la institución financiera	Meses con atraso del titular en los últimos 12 meses	1.5%	3.9%
	Saldo de activo vencido máximo en los últimos 12 meses del titular	2.4%	
Mora sistema financiero	Diferencia entre meses con buen comportamiento y meses con mal comportamiento en el SF en los últimos 24 reportes de la RCC de todos los intervinientes	2.6%	12.2%
	Peor clasificación de riesgo de todos los intervinientes en los últimos 24 reportes de la RCC	9.7%	
Pasivo en la institución financiera	Saldo de pasivo mínimo en los últimos 6 meses del titular	10.8%	10.8%
Solicitante	Zona geográfica	6.7%	22.0%
	Meses transcurridos desde la constitución de la Pyme	15.3%	
Vinculación	Antigüedad máxima del activo o pasivo de los intervinientes en la institución financiera.	3.8%	3.8%

Fuente: Capítulo VI y Elaboración propia

En el anexo 1 se encuentra el detalle de los estimadores de la regresión logística.

6.4. Performance del Credit Scoring Pyme Revolvente

En la tabla VI.6, se aprecia los principales indicadores Gini y Roc. El indicador Gini del modelo tiene un valor de 51.5%, dicho valor indica que el performance del modelo es muy bueno. El indicador Roc toma un valor de 75.7% lo cual confirma la estabilidad del modelo. Los indicadores Gini y Roc miden la capacidad predictiva del modelo.

Tabla VI.6 Principales indicadores Gini - ROC

Asociación de probabilidades predichas y respuestas observadas			
Concordancia de porcentaje	75.5	D de Somers	0.515
Discordancia de porcentaje	24	Gamma	0.517
Porcentaje ligado	0.5	Tau-a	0.108
Pares	26583388	c	0.757

Fuente: Capítulo VI y Elaboración propia

6.5. Tabla de clasificación muestra test Pyme Revolvente

En la tabla VI.7, se aprecia la tasa correcta de clasificación que toma un valor de 79.7% en la base test (fuera de muestra), dicho resultado garantiza la estabilidad del modelo. Para calcular la tasa correcta de clasificación basta sumar los valores donde el resultado observado y el predicho coinciden ($3,961 + 264$) y dividir dicho resultado entre el total de observaciones ($3961+709+369+264$).

Tabla VI.7 Tabla de clasificación muestra test Pyme Revolvente

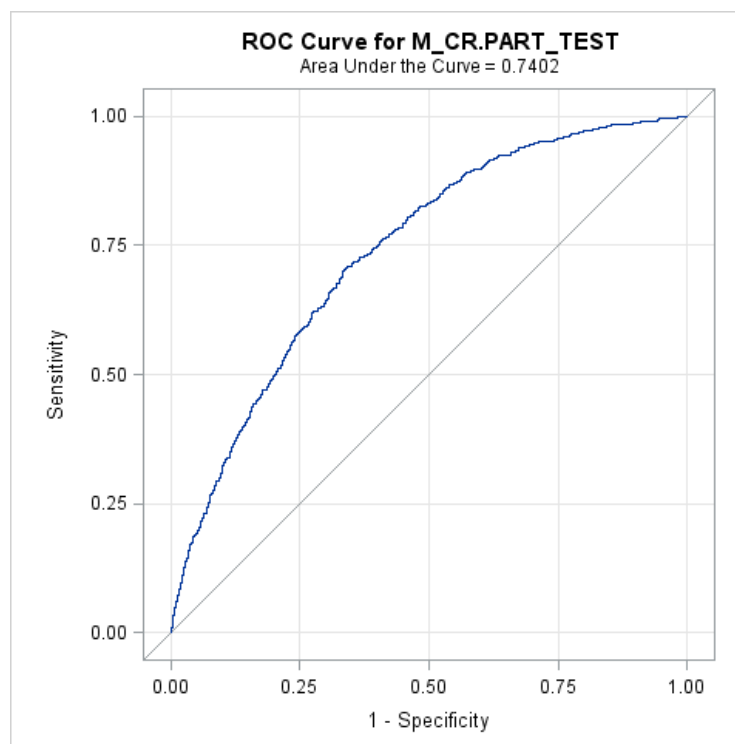
Observado	Predicho	
	0	1
0	3,961	709
1	369	264

Fuente: Capítulo VI y Elaboración propia

6.6. Curva ROC del Credit Scoring Pyme Revolvente en la muestra test

En la figura VI.17, se aprecia que el área bajo la curva toma un valor de 0.742 para la muestra test (fuera de muestra), lo cual garantiza la estabilidad predictiva del modelo.

Figura VI.17 Curva Roc Pyme Revolvente



Fuente: Capítulo VI y Elaboración propia

6.7. Validación del Credit Scoring Revolvente en personas naturales y jurídicas

En este punto vamos a probar si al aplicar el Credit Scoring en el producto Pyme “capital de trabajo” se determina la existencia de diferencias significativas para personas natural o jurídica.

Primero vamos a probar intuitivamente la significancia de dicha variable en el modelo, para ello observamos que modelando con las mismas variables del Credit Scoring Revolvente, podemos correr el modelo solo con las personas naturales y notamos como las variables que presentan mayor relevancia son variables del sistema financiero como por ejemplo el número de empresas que reportan al titular, ratio de evolución de la deuda últimos 3 reportes entre los últimos 24 reportes, máxima clasificación los últimos

24 meses, los saldos pasivos en la institución financiera, los meses de constitución de la empresa y la zona geográfica de la Pyme.

Si corremos el Credit Scoring solo con información de las personas jurídicas las variables que más discriminan en el modelo son los meses con sobregiro en los últimos 6 meses, el patrimonio, los saldos pasivos en la institución financiera y los meses de constitución de la empresa.

En la tabla VI.8, se examina intuitivamente la significancia de la variable tipo de persona natural o jurídica, lo que se refleja en el siguiente cuadro:

Tabla VI.8 Validación del Modelo en personas naturales y jurídicas

Variable	Pesos		
	Credit Scoring Revolvente (Naturales y Jurídicas)	Credit Scoring Revolvente solo con personas Naturales	Credit Scoring Revolvente solo con personas jurídicas
Saldo de activo últimos 3 meses entre saldo activo últimos 24 meses	1.6%	1.7%	2.6%
Número de empresas que reportan al titular	19.6%	16.6%	8.5%
Meses con Sobregiro en los últimos 6 meses	6.6%	2.1%	17.5%
Evolución de la deuda últimos 3 reportes versus los últimos 24 reportes	9.5%	11.4%	7.2%
Pasivo entre patrimonio	1.1%	1.4%	0%
Capital de trabajo neto	2.3%	4.2%	0%
Patrimonio	6.6%	3.7%	15.0%
Atraso máximo en los últimos 12 meses	1.5%	2.4%	0%
Máximo saldo vencido en los últimos 12 meses	2.4%	1.2%	6.8%
Diferencia entre el número de meses con buen y mal comportamiento de los intervinientes en el SF, en los últimos 24 meses.	2.6%	2.0%	3.4%
Máxima clasificación los últimos 24 meses.	9.7%	14.2%	6.8%
Saldo pasivos	10.8%	9.3%	12.9%
Meses de constitución de la empresa	15.3%	16.4%	15.7%
Máxima antigüedad activo o pasivo	3.8%	2.9%	3.4%
Zona geográfica	6.7%	10.4%	0%
Total	100.0%	100.0%	100.0%

Fuente: Capítulo VI y Elaboración propia

El peso en los modelos se calcula a través del estadístico de Wald. Para obtener los valores del estadístico de Wald se calcula la regresión logística en los tres modelos:

- ✓ Credit Scoring revolving (Personas naturales y jurídicas)
- ✓ Credit Scoring revolving (Solo Personas naturales)
- ✓ Credit Scoring revolving (Solo Personas jurídicas)

El valor de los pesos de cada variable, en cada modelo se calcula mediante el estadístico de wald y se define (valor del estadístico de Wald de la variable/suma de valores del estadístico de Wald de todas las variables incluidas en el modelo):

$$Peso_{x_i} = \frac{Wald_{x_i}}{\sum_i^n Wald_{x_i}}$$

Donde i toma valores de 1 hasta n, donde n es el número de variables del modelo.

Finalmente podemos decir que el modelo es estable para personas naturales y jurídicas. Debido a que la variable tipo de persona (Natural o jurídica) es significativa es decir tiene un indicador p-value menor a 0.05 por lo tanto podemos agregarla en el modelo Credit Scoring Revolving.

En el Anexo 2 se agrega la variable tipo de persona (Natural o jurídica) al modelo Credit Scoring Revolving para capital de trabajo.

6.8. Credit Scoring con solo variables del Titular de la Pyme

Una vez generado el modelo Credit Scoring revolving para capital de trabajo, deseamos examinar que pasaría si generamos un modelo, pero con solo variables del titular y algunas variables propias de la Pyme, intuitivamente sabemos que, si el titular de la Pyme es un mal pagador entonces es muy probable que como Pyme también sea un mal pagador. Se genera el modelo con solo variables del titular y se valida que dicho modelo es tan bueno como el modelo revolving construido con información del titular de la Pyme y de los intervinientes (es decir todos los socios de la Pyme). Cuando se dice solo con variables del titular la información que ingresa al modelo contiene exclusivamente información del titular de la Pyme, por ejemplo, los saldos activos o pasivos solo reflejan las deudas o ahorros del titular. En la tabla VI.9, se aprecia el modelo con solo información del titular:

Tabla VI.9 Credit Scoring con solo variables del titular de la Pyme

Modelo con información del titular		ROC =75% GINI:49.9%	
Grupo	Descripción	Peso	Peso del grupo
Endeudamiento en la institución financiera	Ratio de la deuda del titular en la institución financiera el mes previo a la solicitud sobre su deuda máxima en los 24 meses previos a la solicitud	7.5%	7.5%
	Meses con sobregiro en los últimos 6 meses en el sistema financiero del titular	8.9%	28.3%
Información financiera	Número de entidades financieras que reportaron deudas del titular en el mes previo a la solicitud	14.9%	
	Ratio de saldos medios de deuda del titular en el SF de los últimos 6 meses respecto a los 24 últimos meses	4.6%	
	Es el monto total del patrimonio de la empresa.	5.8%	11.4%
	Activo corriente menos pasivo corriente	3.3%	
	Pasivo corriente / patrimonio	2.3%	
Mora en la institución financiera	Peor atraso del titular en la institución financiera en los 3 meses previos a la solicitud	3.5%	5.2%

	Meses con atraso del titular en los últimos 12 meses	1.6%	
Mora Sistema Financiero	Diferencia de meses con buen comportamiento y meses con mal comportamiento en el SF del titular en los 24 meses previos a la solicitud	7.1%	12.5%
	Número de meses transcurridos desde la última vez que se reportó al titular con la clasificación CPP hasta el mes del último reporte RCC en los últimos 12 meses	5.3%	
Pasivo en la institución financiera	Saldo pasivo medio en la institución financiera del titular en los 12 meses previos a la solicitud	7.5%	7.5%
Solicitante	Meses constitución de la empresa	13.5%	23.6%
	Ubicación geográfica basada en direcciones	9.1%	
	Indicador si la empresa es persona natural o jurídica	1.0%	
Vinculación	Meses transcurridos desde la concesión de la operación de activo o pasivo más antigua del titular.	3.9%	3.9%

Fuente: Capítulo VI y Elaboración propia

Se aprecia que el modelo es estable ya que presenta indicadores bastante aceptables en indicador Gini es de 49.9% y un valor ROC de 75%.

Los resultados indican que el riesgo de la Pyme puede ser medido con solo variables del titular y variables solo de la Pyme. El aporte de los intervinientes tiene un peso relativo no muy marcado. Por tanto, si cruzamos en un eje la probabilidad del modelo con variables solo del titular y en el otro eje la probabilidad del Credit Scoring revolvente que tiene variables del titular y los intervinientes, se espera que la relación entre ambos modelos sea fuerte es decir el nivel de correlación sea elevado. En la tabla VI.10, se aprecia que la correlación entre los modelos Credit Scoring revolvente y el modelo con solo información del titular es igual a 0.85. Esto indica que podemos modelar el comportamiento de la Pyme con información solo del titular.

En el anexo 3 se muestra los estimadores del Modelo con información del titular.

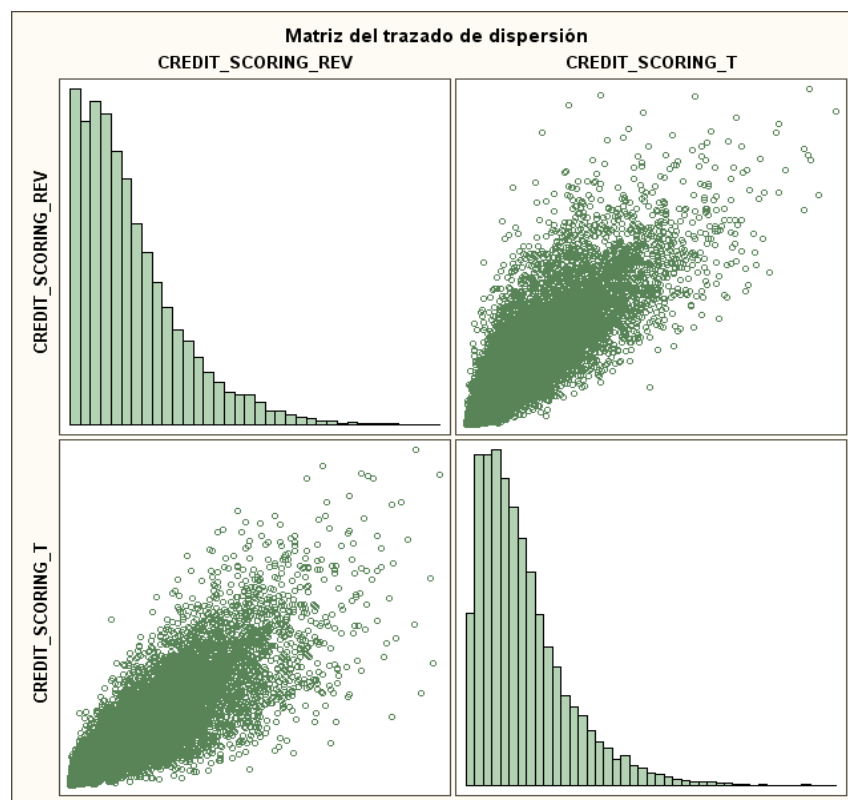
Tabla VI.10 Correlación de Pearson Credit Scoring Revolvente vs Modelo con información del titular

Coeficiente de correlación Pearson, N=15,904		
Prob > r suponiendo H0: Rho=0		
	Credit Scoring Revolvente	Modelo con solo información del titular
Credit Scoring Revolvente	1	0.84633
Modelo con solo información del titular	0.84633	1

Fuente: Capítulo VI y Elaboración propia

En la figura VI. 18 se observa la matriz de correlaciones y las distribuciones del modelo revolvente y el modelo con solo variables del titular.

Figura VI.18 Matriz de Correlaciones Modelo Revolvente vs Modelo Variables del Titular



Fuente: Capítulo VI y Elaboración propia

Se dice que, si es posible modelar la Pyme con información solo del titular, se encuentra evidencia que el riesgo de la Pyme es el riesgo del titular.

6.9. Validación del Modelo

El proceso de validación tiene por objeto comprobar la discriminación que proporciona el modelo para diferentes subconjuntos de la población original.

6.9.1 Indicadores de Bondad de ajuste del Credit Scoring Revolvente

En la tabla VI.11, se aprecia el resumen de los indicadores de discriminación del modelo revolvente sobre las muestras train (intramuestral) y test (fuera de muestra).

Tabla VI.11 Indicadores de Bondad de Ajuste

Segmento	Muestra	ROC	KS	Gini
Revolvente	Train	77.5%	42.4%	51.5%
	Test	74.0%	39.4%	48.0%

Fuente: Capítulo VI y Elaboración propia

El performance obtenido por el modelo presenta un indicador Gini de 51.5% para la muestra de construcción train (intramuestral) y de 48% para la muestra test (fuera de muestra). El nivel de predictivo es elevado en ambos casos independiente de la muestra.

6.9.2 Intervalos de confianza para el Credit Scoring Revolvente

La metodología bootstrap se ha aplicado al modelo revolvente, para ello se realizó los siguientes pasos:

1. El tamaño de cada réplica es igual al 100% del total de la base de entrenamiento train (intramuestral) del modelo y también para la muestra test (fuera de muestra).
2. El número de réplicas es igual a 4000 muestras en ambas muestras train (intramuestral) y test (fuera de muestra).
3. Con estos parámetros se construyeron los intervalos de confianza de 95% y 99% para las muestras train (intramuestral) y test (fuera de muestra).

En la tabla VI.12, se aprecia los intervalos de confianza al 95% y al 99% para el Gini en la muestra train (intramuestral) y test (fuera de muestra). Por ejemplo, podemos decir que el 95% de las veces el Gini de la muestra train está en el intervalo [49.4% - 53.7%].

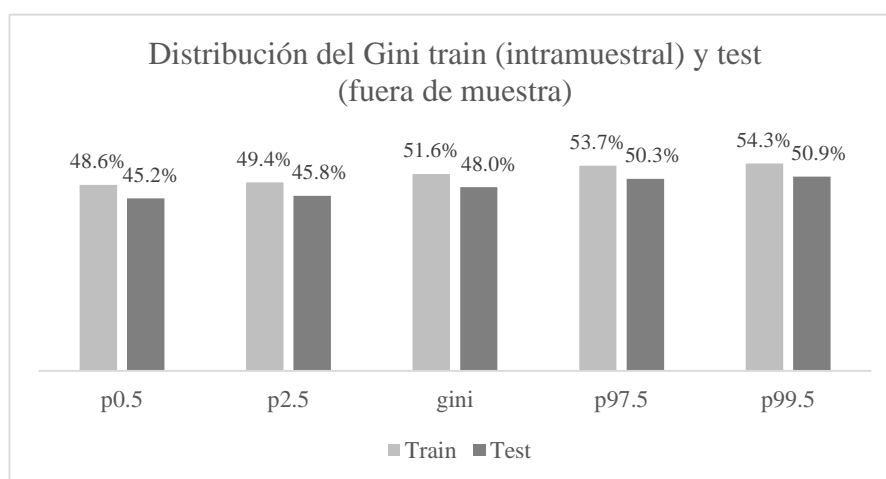
Tabla VI.12 Intervalos de confianza Credit Scoring revolving

Muestra	Total	Default	% Default	P0.5	P2.5	Gini	P97.5	P99.5
Train (intra-muestral)	15,904	1898	11.93%	48.6%	49.4%	51.5%	53.7%	54.3%
Test (fuera de muestra)	5,303	633	11.94%	45.2%	45.8%	48.0%	50.3%	50.9%

Fuente: Capítulo VI y Elaboración propia

En la figura VI.19, se aprecia los intervalos de confianza al 95% y al 99% para el Gini en la muestra de entrenamiento train (intramuestral) y test (fuera de muestra). Los percentiles de la muestra test siempre son menores que la muestra de entrenamiento, se concluye que hay estabilidad en el indicador de predicción Gini, ya que siempre está dentro de los límites de predicción al 95% y al 99% de los casos.

**Figura VI.19 Intervalo de confianza muestra train y test
Credit Scoring Revolvente**



Fuente: Capítulo VI y Elaboración propia

Si el Gini del modelo en la etapa de monitoreo disminuye por debajo de 45.8% (percentil 2.5) de la muestra test entonces sería necesario calibrar el modelo.

Se concluye que el modelo es estable en ambas muestras train (intramuestral) y test (fuera de muestra).

CAPÍTULO VII. CONCLUSIONES

En la presente tesis se ha revisado las principales metodologías de Credit Scoring, basadas en modelos Paramétricos, No Paramétricos y Semi Paramétricos.

Entre los modelos Paramétricos se aprecia:

- ✓ Análisis discriminante tiene un buen rendimiento en grandes muestras, pero estimadores ineficientes.
- ✓ El modelo Z- score es una combinación lineal de ratios financieros para predecir quiebra de empresas, presenta estimaciones robustas, pero el modelo solo aplica a personas jurídicas y no a personas naturales.
- ✓ El modelo Least – Absolute – Value (LAV), minimiza la suma de los valores absolutos de los residuos y las estimaciones son robustas, es similar a los mínimos cuadrados ordinarios, presenta múltiples soluciones y frente a valores atípicos las estimaciones son inestables.
- ✓ Modelos Lineales probabilísticos (MLP), si es una combinación lineal de variables óptimas puede predecir una respuesta bivariada eficiente. Sin embargo, pierde precisión si se violan los supuestos de linealidad.
- ✓ Modelo logístico (logit), presenta gran rendimiento, buenas propiedades estadísticas. Corrige la limitación de los modelos lineales probabilísticos (MLP). Fácil interpretación de los datos y es usado en la banca para gestionar el riesgo de crédito.
- ✓ Modelo Probabilístico (Probit), buen rendimiento, normalidad en la predicción, mayor estabilidad en las variables. Sin embargo, es de difícil interpretación en la data.
- ✓ Modelo Tobit, proporcionan un mejor ajuste a la realidad de los datos, pero su uso es menos comercial por su nivel de complejidad.
- ✓ Modelos logit multinomiales, generalizan la regresión logística con más de dos posibles resultados discretos, su aprendizaje es lento si las categorías a predecir son muchas. Se enfrenta a algoritmos que aprenden rápido como Naive Bayes.

- ✓ Modelos logit mixtos, no tiene las limitaciones del logit estándar, variaciones estándar de la muestra, permite utilizar parámetros simulados. Presentan estimación compleja de parámetros mediante simulación de Monte Carlo y generan mayor error tipo I.

Entre los modelos No Paramétricos revisados se aprecia que:

- ✓ Las Redes neuronales, presentan gran predicción y flexibilidad, pero son una caja negra de difícil interpretación.
- ✓ Los modelos Support Vector Machine (SMV) presentan gran predicción, pero son difíciles de interpretar, poco intuitivos.
- ✓ Modelos bayesianos (Naive Models) requieren de una muestra pequeña de datos, se limitan a las características que ingresan al modelo.

Entre los modelos Semi paramétricos revisados se aprecia que:

- ✓ Máximum Score, se obtienen estimadores robustos de los parámetros del modelo. Sin embargo, presenta problemas de heterocedasticidad y correlación.
- ✓ Técnica Chaid (Chi Squared Automatic Interaction Detection), los resultados son altamente visuales y de fácil interpretación, pero necesita de tamaños de muestra bastante grandes por la cantidad de divisiones que realiza.
- ✓ LS – SVM (Least Squares Support Vector Machine), utiliza kernel o núcleos y son estadísticamente estables, pero son de difícil interpretación.
- ✓ Neuronal Network más Logit Models se aplica varias técnicas como componentes principales, algoritmo genético en la red neuronal de kohonen. Son de elevado esfuerzo computacional.
- ✓ Decisión tree fácil de entender, requiere poca preparación de los datos, visual, eficaz para predecir y clasificar. Su limitación es que tiene problemas de sobreajuste.

En la presente tesis se aplica el modelo logístico ya que es usado por las instituciones financieras por su fácil interpretación, bondad de ajuste matemático y sobre todo porque establece una línea de comparación con las diferentes metodologías presentadas y porque es un modelo sólido en sus parámetros estadísticos. No hay una metodología óptima de Credit Scoring, mucho depende de la calidad de la información y de las variables que ingresan al modelo.

El presente trabajo tiene como objetivo principal estimar la probabilidad de incumplimiento de pago de un crédito Pyme, en este caso la aplicación se ha realizado sobre el producto capital de trabajo en la modalidad revolvente, por la demanda que presenta dicho producto. Se obtiene un indicador de discriminación Gini aceptable de 51.5% en la base de construcción (intramuestral) y de 48% en la base de test (fuera de muestra). La metodología elegida fue la regresión logística binaria por su ventaja en el cálculo de la probabilidad, facilidad de interpretación, rápida ejecución computacional y es aplicada para la gestión del riesgo de crédito en el sistema financiero peruano. El modelo es de admisión de clientes, es decir es muy preciso en la aprobación de créditos y se aplica solo a personas que tienen una experiencia previa en el sistema financiero. El modelo contiene información del dueño de la Pyme y de los intervinientes y se diferencia de los demás modelos porque no es un modelo genérico, porque está construido en función a la data interna de la institución financiera, no es un modelo de buró genérico, tampoco utiliza información de otros scoring como input, se aplica a un segmento específico Pyme revolvente, se ha visto en la revisión bibliográfica que muchos de estos modelos utilizan además de variables internas y de buró externos o de centrales de riesgo. El modelo analiza una profundidad histórica de 24 meses, detalle que no se observa en ninguno de los modelos revisados. La ventana de predicción es de un año lo que implica estimaciones estables.

Un aporte relevante del modelo es la estabilidad probada intramuestral y fuera de muestra ya que en este sentido se ha calculado 4,000 simulaciones de Monte Carlo para el indicador GINI y ROC determinando que dichos indicadores están dentro de los intervalos de confianza al 95% y 99%. Por lo tanto, el modelo propuesto nos garantiza un alto rendimiento predictivo ya que ha sido sometido a pruebas de estabilidad predictiva (Simulación de Monte Carlo- Bootstrap). En ninguna de las aplicaciones revisadas se

brinda un intervalo de confianza para sus estimadores de predicción intramuestral y fuera de muestra.

Se evalúa las principales características o variables que influyen en el comportamiento de pago, determinando los siguientes grupos de variables:

- a) Las variables del grupo Endeudamiento en la institución financiera, tienen un peso como grupo de 1.6%:
 - ✓ Se aprecia que si aumenta el saldo activo de los últimos tres meses entre saldo de activo los últimos 24 meses entonces la probabilidad de default aumenta. La variable tiene un nivel de importancia o peso de 1.6% y es significativa.
- b) Las variables del grupo Endeudamiento en el sistema financiero, tienen un peso como grupo de 35.6%:
 - ✓ Se aprecia que, si aumenta el número de empresas que reportan al titular de la Pyme entonces aumenta el riesgo de incumplimiento. La variable tiene un nivel de importancia o peso de 19.6% y es significativa.
 - ✓ Se aprecia que, si aumenta la cantidad de meses con sobregiro entonces la tasa de incumplimiento se incrementa. La variable tiene un nivel de importancia o peso de 6.6% y es significativa.
 - ✓ Se aprecia que, si la Pyme aumenta el nivel de endeudamiento entonces el riesgo de incumplimiento crece. La variable tiene un nivel de importancia o peso de 9.5% y es significativa.
- c) Las variables del grupo información financiera, tienen un peso como grupo de 10%:
 - ✓ Se aprecia que, si la Pyme toma mayor apalancamiento el riesgo de incumplimiento aumenta. La variable tiene un nivel de importancia o peso de 1.1% y es significativa.
 - ✓ Se aprecia que, si aumenta el capital de trabajo neto, es decir la liquidez de la Pyme entonces la tasa de incumplimiento disminuye. La variable tiene un nivel de importancia o peso de 2.3% y es significativa.

- ✓ Se aprecia que, si aumenta el patrimonio entonces el riesgo disminuye. La variable tiene un nivel de importancia o peso de 6.6% y es significativa.
- d) Las variables del grupo mora en la institución financiera, tienen un peso como grupo de 3.9%:
- ✓ Se aprecia que, si aumenta el atraso entonces el riesgo aumenta. La variable tiene un nivel de importancia o peso de 1.5% y es significativa.
 - ✓ Se aprecia que, si aumenta el saldo activo vencido entonces el riesgo aumenta. La variable tiene un nivel de importancia o peso de 2.4% y es significativa.
- e) Las variables del grupo mora en el sistema financiero, tienen un peso como grupo de 12.2%:
- ✓ Se aprecia que, si aumenta la diferencia entre meses con buen y mal comportamiento entonces la tasa de incumplimiento disminuye. La variable tiene un nivel de importancia o peso de 2.6% y es significativa.
 - ✓ Se aprecia que, si la Pyme tiene peor clasificación en sistema financiero entonces la tasa de incumplimiento aumenta. La variable tiene un nivel de importancia o peso de 9.7% y es significativa.
- f) Las variables de pasivo en la institución financiera, tienen un peso como grupo de 10.8%:
- ✓ Se aprecia que, si aumentan los pasivos de la Pyme entonces la tasa de incumplimiento disminuye. La variable tiene un nivel de importancia o peso de 10.8% y es significativa.
- g) Las variables del grupo solicitante, tienen un peso como grupo de 22%:
- ✓ Se aprecia que, si las zonas son de mayor riesgo, como por ejemplo el Callao entonces la tasa de incumplimiento aumenta. La variable tiene un nivel de importancia o peso de 6.7% y es significativa.

- ✓ Se aprecia que, si aumentan los meses de constitución de la Pyme entonces el riesgo de incumplimiento disminuye. La variable tiene un nivel de importancia o peso de 15.3% y es significativa.

h) Las variables del grupo de vinculación, tienen un peso como grupo de 3.8%:

- ✓ Se aprecia que, si aumenta la antigüedad de la Pyme en la institución financiera entonces el riesgo de incumplimiento disminuye. La variable tiene un nivel de importancia o peso de 3.8% y es significativa.

En resumen las características o variables más importantes para el Credit Scoring que ayudan a explicar una disminución o aumento del riesgo de crédito son: (i) el saldo activo de los últimos tres meses entre el saldo activo de los últimos 24 meses, (ii) el número de empresas que reportan al titular, (iii) los sobregiros, (iv) la evolución de la deuda últimos tres reportes entre los últimos 24 meses, (v) el pasivo entre el patrimonio, (vi) el capital de trabajo, (vii) el patrimonio, (viii) el atraso máximo en los últimos 12 meses, (ix) el máximo saldo vencido en los últimos 12 meses, (x) la diferencia de meses con buen comportamiento y mal comportamiento, (xi) la máxima clasificación en el sistema financiero los últimos 24 meses, (xii) los pasivos, (xiii) los meses de constitución de la empresa, (xiv) la máxima antigüedad en pasivos o activos y (xv) la zona geográfica. En el modelo todas las variables son significativas con un p-value menor a 0.05 y tienen sentido económico y de negocio.

Se aprecia en el modelo Credit Scoring que existen diferencias significativas entre persona natural y jurídica. El modelo en mención se aplica solo a personas naturales y se observa que las variables que toman mayor peso o importancia son las variables de la experiencia en el sistema financiero (empresas donde reportan al titular, evolución de deudas y clasificación), saldos pasivos, meses de constitución de la Pyme y la zona geográfica. Cuando se aplica el modelo a solo personas jurídicas el peso de las variables se inclina al sobregiro, patrimonio, saldos pasivos y meses de constitución de la Pyme. Las diferencias existentes entre los grupos de personas naturales y jurídicas son confirmadas por el modelo agregado, en el que la variable ficticia que hace referencia a los grupos (natural y jurídica) es estadísticamente significativa. Finalmente se considera la

variable “persona natural jurídica” en el modelo como una variable adicional y se observa que es estadísticamente significativa con un p-value menor a 0.05 e incrementa el indicador Gini a 51,6%.

Uno de los aportes relevantes de la tesis es que se verifica que el riesgo de la Pyme es el riesgo del titular, se genera un modelo con solo variables del titular y se observa que su indicador Gini es de 49.9%, comparándolo con el modelo Credit Scoring revolvente (con información del titular y los socios) que tiene un Gini de 51.5% se observa que la diferencia no es marcada, lo que intuitivamente nos indica que el aporte de los socios o intervinientes no es relevante. Esto quiere decir que podemos generar un modelo para Pyme con información solo del dueño de la Pyme.

Ambos modelos el Credit Scoring con solo variables del titular y el Credit Scoring revolvente presentan alto poder de predicción y tienen valores cercanos. Gráficamente se ponen ambos modelos en un gráfico de dispersión se aprecia que su correlación es 0.85 eso confirma que el riesgo de Pyme está correlacionado con el riesgo del titular de la Pyme y según el grafico de correlación la relación entre los modelos es más estable cuando las probabilidades de incumplimiento de ambos modelos no son elevadas.

Se encuentra evidencia que el riesgo de la Pyme es el riesgo del titular. Sin embargo, se sugiere profundizar en este aspecto en estudios posteriores.

CAPÍTULO VIII. RECOMENDACIONES

Se deben incorporar y probar a través de pilotos, nuevas metodologías de Credit Scoring aplicados al segmento Pyme, por consiguiente, se reten a las metodologías presentadas. Se podrían adicionar técnicas deep learning, xgboost, entre otras.

Se debe garantizar la estabilidad del modelo en el tiempo, mediante un seguimiento periódico de sus variables y de la probabilidad calculada. Se recomienda una revisión como mínimo una vez al año.

Se sugiere en futuros trabajos agregar más variables como por ejemplo edad, sexo, estado civil, garantías reales, tipo de cambio, entre otras. Cabe mencionar que dicha información no se encontraba disponible en la base de datos de la institución financiera.

Se sugiere crear estrategias diferenciadas para personas naturales y jurídicas, ya que el modelo nos indica que hay diferencias significativas en estos grupos.

Profundizar el análisis sobre la medición del riesgo de crédito con solo la información del titular de la Pyme, se podrían agregar más variables del tipo cualitativo que consoliden la prueba que el riesgo del titular es el riesgo de la Pyme. Se sugiere probar si el porcentaje de participación a nivel de acciones discrimina o no el riesgo de crédito en la Pyme, entre otras pruebas.

ANEXOS

ANEXO 1: CREDIT SCORING CAPITAL DE TRABAJO REVOLVENTE

Número de observaciones leí	15904
Número de observaciones usa	15904

Perfil de respuesta		
Valor ordenado	FLG_MORA60	Frecuencia total
1	1	1898
2	0	14006

La probabilidad modelada es FLG_MORA60=1

Tipo 3 Análisis de efectos			
Efecto	DF	Chi-cuadrado de Wald	Pr > ChiSq
RATIO_ACT_SOL_3_24_Tv2	1	11.1581	0.0008
CTDEMPREPORTADOCLIMED1	1	134.8839	<.0001
MSG6	1	45.1294	<.0001
SF3_SF24_V2_T	1	65.2736	<.0001
PASIVO_PATRIMONIO	1	7.4675	0.0063
CAPITAL_TRABAJO_NETO	1	15.7752	<.0001
MTOTOTALPATRIMONIOSOL	1	45.472	<.0001
ATRASO_MESES12	1	10.1492	0.0014
SALDO_VENCIDO_MAX12	1	16.7899	<.0001
MESES_ACTIVADO_SF_BU_MA24_T	1	17.6574	<.0001
TIPCLASIFRIESGOSBSMAX24_T	1	66.5307	<.0001
MTOSALDOPROTOTAL_SOL_PASMIN6	1	74.1603	<.0001
MESES_CONSTITUCION_P	1	105.5329	<.0001
maxantigüedad_act_pas_tot_T	1	26.386	<.0001
MACROZONA_GZ_T_P	12	45.8183	<.0001

Análisis del estimador de máxima verosimilitud					
Parámetro	DF	Estima- dor	Error estándar	Chi-cua- drado de Wald	Pr > ChiSq
Intercept	1	-1.9307	0.0915	445.668	<.0001
RATIO_ACT_SOL_3_24_Tv2	1	0.051	0.0153	11.1581	0.0008
CTDEMPREPORTADOCLIMED1	1	0.206	0.0177	134.884	<.0001
MSG6	1	0.1155	0.0172	45.1294	<.0001
SF3_SF24_V2_T	1	0.1391	0.0172	65.2736	<.0001
PASIVO_PATRIMONIO	1	0.0712	0.026	7.4675	0.0063
CAPITAL_TRABAJO_NETO	1	-2.10E-06	5.30E-07	15.7752	<.0001
MTOTOTALPATRIMONIOSOL	1	-1.91E-06	2.83E-07	45.472	<.0001
ATRASO_MESES12	1	0.0511	0.016	10.1492	0.0014
SALDO_VENCIDO_MAX12	1	0.000154	0.000038	16.7899	<.0001
MESES_ACTIVADO_SF_BU_MA24_T	1	-0.014	0.00334	17.6574	<.0001
TIPCLASIFRIESGOSBSMAX24_T	1	0.196	0.024	66.5307	<.0001
MTOSALDOPROTOTAL_SOL_PAS- MIN6	1	-0.00007	8.34E-06	74.1603	<.0001
MESES_CONSTITUCION_P	1	-0.00539	0.000525	105.533	<.0001
maxantigüedad_act_pas_tot_T	1	-0.00303	0.000591	26.386	<.0001
MACROZONA_GZ_T_P	1	-0.1284	0.0911	1.9883	0.1585
MACROZONA_GZ_T_P	2	0.0497	0.0877	0.3212	0.5709
MACROZONA_GZ_T_P	4	-0.1409	0.1677	0.7062	0.4007
MACROZONA_GZ_T_P	5	-0.1501	0.0695	4.6667	0.0308
MACROZONA_GZ_T_P	6	0.0149	0.1079	0.0192	0.8899
MACROZONA_GZ_T_P	8	0.3566	0.0716	24.769	<.0001
MACROZONA_GZ_T_P	C	-0.0352	0.153	0.0529	0.8181
MACROZONA_GZ_T_P	LC	-0.1892	0.0727	6.7811	0.0092
MACROZONA_GZ_T_P	LE	-0.0753	0.0969	0.6047	0.4368
MACROZONA_GZ_T_P	LM	-0.0223	0.0899	0.0617	0.8039
MACROZONA_GZ_T_P	LN	-0.046	0.0867	0.2808	0.5962
MACROZONA_GZ_T_P	LP	0.2803	0.1331	4.4339	0.0352

Asociación de probabilidades predichas y respuestas observadas			
Concordancia de porcentaje	75.5	D de Somers	0.515
Discordancia de porcentaje	24	Gamma	0.517
Porcentaje ligado	0.5	Tau-a	0.108
Pares	26583388	c	0.757

ANEXO 2 CREDIT SCORING CAPITAL DE TRABAJO REVOLVENTE ADICIONANDO LA VARIABLE PERSONA NATURAL O JURIDICA

Número de observaciones leí	15904
Número de observaciones usa	15904

Perfil de respuesta		
Valor ordenado	FLG_MORA60	Frecuencia total
1	1	1898
2	0	14006

La probabilidad modelada es FLG_MORA60=1

Probar hipótesis nula global: BETA=0			
Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosím	1428.9795	27	<.0001
Puntuación	1225.6987	27	<.0001
Wald	1045.3313	27	<.0001

Tipo 3 Análisis de efectos			
Efecto	DF	Chi-cuadrado de Wald	Pr > ChiSq
RATIO_ACT_SOL_3_24_Tv2	1	12.7376	0.0004
CTDEMPREPORTADOCLIMED1	1	90.0871	<.0001
MSG6	1	52.4834	<.0001
SF3_SF24_V2_T	1	64.1828	<.0001
PASIVO_PATRIMONIO	1	8.5431	0.0035
CAPITAL_TRABAJO_NETO	1	14.4109	0.0001
MTOTOTALPATRIMONIOSOL	1	38.8017	<.0001
ATRASO_MESES12	1	9.775	0.0018
SALDO_VENCIDO_MAX12	1	16.8851	<.0001
MESES_ACTIVO_SF_BU_MA24_T	1	15.5531	<.0001
TIPCLASIFRIESGOSBSMAX24_T	1	71.6994	<.0001
MTOSALDOPROTOTAL_SOL_PASMIN6	1	70.5615	<.0001
MESES_CONSTITUCION_P	1	112.3282	<.0001
maxantigüedad_act_pas_tot_T	1	22.2426	<.0001
MACROZONA_GZ_T_P	12	42.8172	<.0001
INDNATURALJURIDICO	1	7.6485	0.0057

Análisis del estimador de máxima verosimilitud					
Parámetro	DF	Estimador	Error estándar	Chi-cuadrado de Wald	Pr > ChiSq
Intercept	1	-1.9681	0.0923	454.8557	<.0001
RATIO_ACT_SOL_3_24_Tv2	1	0.0548	0.0153	12.7376	0.0004
CTDEMPREPORTADOCLIMED1	1	0.1842	0.0194	90.0871	<.0001
MSG6	1	0.1403	0.0194	52.4834	<.0001
SF3_SF24_V2_T	1	0.1378	0.0172	64.1828	<.0001
PASIVO_PATRIMONIO	1	0.0769	0.0263	8.5431	0.0035
CAPITAL_TRABAJO_NETO	1	-2.01E-06	5.30E-07	14.4109	0.0001
MTOTOTALPATRIMONIOSOL	1	-1.77E-06	2.85E-07	38.8017	<.0001
ATRASO_MESES12	1	0.0502	0.016	9.775	0.0018
SALDO_VENCIDO_MAX12	1	0.000155	0.000038	16.8851	<.0001
MESES_ACTIVO_SF_BU_MA24_T	1	-0.0132	0.00336	15.5531	<.0001
TIPCLASIFRIESGOSBSMAX24_T	1	0.205	0.0242	71.6994	<.0001
MTOSALDOPROTOTAL_SOL_PASMIN6	1	-0.00007	8.24E-06	70.5615	<.0001
MESES_CONSTITUCION_P	1	-0.00565	0.000533	112.3282	<.0001
maxantiguedad_act_pas_tot_T	1	-0.00281	0.000595	22.2426	<.0001
MACROZONA_GZ_T_P	1	-0.1265	0.0911	1.9272	0.1651
MACROZONA_GZ_T_P	2	0.0362	0.0879	0.1696	0.6805
MACROZONA_GZ_T_P	4	-0.1537	0.1678	0.8386	0.3598
MACROZONA_GZ_T_P	5	-0.153	0.0695	4.8487	0.0277
MACROZONA_GZ_T_P	6	-0.00363	0.1081	0.0011	0.9732
MACROZONA_GZ_T_P	8	0.343	0.0718	22.8087	<.0001
MACROZONA_GZ_T_P	C	-0.017	0.1534	0.0124	0.9115
MACROZONA_GZ_T_P	LC	-0.1845	0.0727	6.4364	0.0112
MACROZONA_GZ_T_P	LE	-0.0678	0.097	0.4882	0.4847
MACROZONA_GZ_T_P	LM	0.00619	0.0906	0.0047	0.9455
MACROZONA_GZ_T_P	LN	-0.0329	0.0869	0.1431	0.7052
MACROZONA_GZ_T_P	LP	0.2647	0.1333	3.9445	0.047
INDNATURALJURIDICO	J	-0.1019	0.0369	7.6485	0.0057

Asociación de probabilidades predichas y respuestas observadas			
Concordancia de porcentaje	75.5	D de Somers	0.516
Discordancia de porcentaje	24	Gamma	0.518
Porcentaje ligado	0.5	Tau-a	0.108
Pares	3E+07	c	0.758

ANEXO 3 CREDIT SCORING REVOLVENTE CON SOLO VARIABLES DEL TITULAR

Número de observaciones leí	32115
Número de observaciones usa	32115

Perfil de respuesta		
Valor ordenado	FLG_MORA60	Frecuencia total
1	1	3138
2	0	28977

La probabilidad modelada es FLG_MORA60=1

Probar hipótesis nula global: BETA=0			
Test	Chi-cuadrado	DF	Pr > ChiSq
Ratio de verosim	2309.1526	27	<.0001
Puntuación	2034.1899	27	<.0001
Wald	1778.3675	27	<.0001

Tipo 3 Análisis de efectos			
Efecto	DF	Chi-cuadrado de Wald	Pr > ChiSq
MONTOADE_ACT_MAX24	1	77.4773	<.0001
MSG6	1	91.1947	<.0001
CTDEMPREPORTADOCLIMED1	1	153.4306	<.0001
SF6_SF24_V2	1	47.1844	<.0001
MTOTOTALPATRIMONIOSOL	1	60.2051	<.0001
CAPITAL_TRABAJO_NETO	1	34.4083	<.0001
PASIVO_PATRIMONIO	1	23.2022	<.0001
CTDDIAATRASOMAX3	1	36.5593	<.0001
ATRASO_MESES12	1	16.8754	<.0001
MESES_ACTIVO_SF_BU_MA24	1	73.6726	<.0001
SF_DIS_CAL_CPP12	1	54.8986	<.0001
MTOSALDOPROTOTAL_SOL_PASMED12	1	77.6228	<.0001
MESES_CONSTITUCION_P	1	138.7534	<.0001
maxantigüedad_act_pas_tot	1	40.4611	<.0001
MACROZONA_GZ_T_P	12	94.0577	<.0001
INDNATURALJURIDICO	1	10.4258	0.0012

Análisis del estimador de máxima verosimilitud					
Parámetro	DF	Estimador	Error estándar	Chi-cuadrado de Wald	Pr > ChiSq
Intercept	1	-1.8502	0.0895	426.92	<.0001
MONTOADE_ACT_MAX24	1	0.4889	0.0555	77.477	<.0001
MSG6	1	0.1372	0.0144	91.195	<.0001
CTDEMPREPORTADOCLIMED1	1	0.1972	0.0159	153.43	<.0001
SF6_SF24_V2	1	0.1457	0.0212	47.184	<.0001
MTOTOTALPATRIMONIOSOL	1	-1.69E-06	2.18E-07	60.205	<.0001
CAPITAL_TRABAJO_NETO	1	-2.34E-06	3.99E-07	34.408	<.0001
PASIVO_PATRIMONIO	1	0.0769	0.016	23.202	<.0001
CTDDIAATRASOMAX3	1	0.0297	0.00492	36.559	<.0001
ATRASO_MESES12	1	0.0457	0.0111	16.875	<.0001
MESES_ACTIVADO_SF_BU_MA24	1	-0.0245	0.00286	73.673	<.0001
SF_DIS_CAL_CPP12	1	-0.035	0.00472	54.899	<.0001
MTOSALDOPROTOTAL_SOL_PAS-MED12	1	-0.00003	3.01E-06	77.623	<.0001
MESES_CONSTITUCION_P	1	-0.0049	0.000416	138.75	<.0001
maxantiguedad_act_pas_tot	1	-0.00445	0.000699	40.461	<.0001
MACROZONA_GZ_T_P	1	-0.2194	0.073	9.0425	0.0026
MACROZONA_GZ_T_P	2	0.0441	0.0677	0.4234	0.5152
MACROZONA_GZ_T_P	4	-0.0873	0.1303	0.4487	0.5029
MACROZONA_GZ_T_P	5	-0.0937	0.0516	3.3058	0.069
MACROZONA_GZ_T_P	6	0.0267	0.0801	0.1112	0.7387
MACROZONA_GZ_T_P	8	0.3348	0.0524	40.867	<.0001
MACROZONA_GZ_T_P	C	0.014	0.1129	0.0155	0.901
MACROZONA_GZ_T_P	LC	-0.271	0.0553	23.984	<.0001
MACROZONA_GZ_T_P	LE	-0.1307	0.076	2.9533	0.0857
MACROZONA_GZ_T_P	LM	-0.0177	0.0774	0.0523	0.8191
MACROZONA_GZ_T_P	LN	-0.00322	0.066	0.0024	0.961
MACROZONA_GZ_T_P	LP	0.2845	0.0914	9.6981	0.0018
INDNATURALJURIDICO	J	-0.0939	0.0291	10.426	0.0012

Asociación de probabilidades predichas y respuestas observadas			
Concordancia de porcentaje	74.7	D de Somers	0.499
Discordancia de porcentaje	24.7	Gamma	0.502
Porcentaje ligado	0.6	Tau-a	0.088
Pares	9.1E+07	c	0.75

BIBLIOGRAFÍA

- Abdou, H. A. & Pointon, J. (2011). Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature. *Intelligent Systems in Accounting, Finance and Management*, 18(2–3), 59–88. <http://doi.org/10.1002/isaf.325>
- Altman, E. & Sabato, G., (2007). Modelling Credit risk for SMEs: Evidence from the US Market. doi: 10.1111/j.1467-6281.2007.00234.x
- Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the prediction of Corporate Bankruptcy. *Journal of finance*, 23(4), 589-609.
- Banasik, J., Crook, J., & Thomas, L. (2003). Sample selection bias in Credit Scoring models. University of Edinburgh. Edinburgh, UK; University of Southampton, Southanmpton, UK. *Journal of the Operational Research Society* (2003) 54, 822–832. doi:10.1057/palgrave.jors.2601578
- Banco Central de Reserva del Perú. (Mayo, 2018). Reporte de Estabilidad Financiera. Recuperado de <http://www.bcrp.gob.pe/publicaciones/reportes-de-estabilidad-financiera/ref-mayo-2018.html>
- Basel Committee on Banking Supervision. (2001). The Internal Ratings-Based Approach (Consultative Document). Basilea, Suiza: Bank for International Settlements- BIS.
- Bellotti, T., & Crook, J. (2007). Credit Scoring with Macroeconomic Variables Using Survival Analysis. Credit Research Centre-Management School and Economics. Edinburgh, UK:University of Edinburgh.
- Bellotti, T., & Crook, J. (2009). Support vector machines for Credit Scoring and discovery of significant features. Credit Research Centre, Management School and Economics, University of Edinburgh, William Robertson Building, 50 George Square, Edinburgh EH8 9JY, UK.
- BIS (2006) La mejora del gobierno corporativo en organizaciones bancarias. Recuperado de <https://www.bis.org/publ/bcbs122es.pdf>

- BIS (2017). Resumen de las reformas de Basilea III. Recuperado de https://www.bis.org/bcbs/publ/d424_hlsummary_es.pdf
- Chumacero Sánchez, A. & Sandoval Lozano, K. (2011). Impacto Financiero en la implementación de un modelo Credit Scoring en una entidad de micro finanzas. (Tesis de maestría finanzas). Universidad Esan, Lima, Perú.
- Coffman, J.Y. (1986). The Proper Role of Tree Analysis in Forecasting The Risk Behaviour of Borrowers. Management Decision Systems , MDS Reports 3,4,7 & 9.
- David Lawrence and Arlene Solomon Managing a Consumer Lending Business – Primera Edición 2002 - New York.
- Davis, R. H., Edelman, D. B., & Gamberman, A. J. (1992). Machine-Learning Algorithms for Credit-Card Applications. Journal of Management Mathematics, 4(1), 43-51.
- DENNIS, W. (1995). Fair lending and Credit Scoring. Mortgage Banking, 56 (2), 55-59
- Diallo, B. (2006). Un modele de Credit Scoring pour une institution de micro-finance Africane: le cas de Nyesigiso au Mali. Recuperado de http://hal.archives-ouvertes.fr/docs/00/06/91/63/PDF/s16_05_06diallo.pdf
- Esquivel Gómez, J. , León Ramirez, M. & Arley Vengas, R. (I semestre, 2013). Modelo de Credit Scoring Pymes para la medición del riesgo de morosidad de pequeña y mediana empresa: Un caso de aplicación en entidad financiera. Recuperado de https://repositoriotec.tec.ac.cr/bitstream/handle/2238/7167/modelo_Credit_Scoring.pdf?sequence=1&isAllowed=y
- Fisher R. A. (1936) The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics, 7(2), 179-188.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7(2), 179–188. <http://doi.org/10.1111/j.1469-1809.1936.tb02137.x>

- Glennon, D.; Kiefer, N. M.; Larson, C. E. & Choi, H. (2008). Development and Validation of Credit Scoring Models. *Journal of Credit Risk*, Forthcoming, 1(1), 1–70. Retrieved from <http://papers.ssrn.com/abstract=1180302>
- Greene, W. H. (1992). A Statistical Model for Credit Scoring (Working Papers 92-29). New York: Leonard N. Stern School of Business, New York University.
- Gutiérrez Girault, M. (Octubre, 2007). Modelos de Credit Scoring - Qué, Cómo, Cuándo y Para qué-. Recuperado de <http://www2.bcra.gob.ar/Pdfs/Publicaciones/CreditScoring.pdf>
- Hand, D. J. (1981). *Discrimination and Classification*. Chichester, UK: Wiley.
- Hand, D.J., & Henley, W. E. (1997). Statistical Classification Methods in Costumer Credit Scoring: A review. *Journal of the Royal Statistical Association*, 160(A/Part3), 523-541.
- Hidalgo Montalvo, M., & Pichardo Sihuinchá, E. (2015). Score de Bureau para cajas municipales de ahorro y créditos. (Tesis de maestría en finanzas). Universidad Esan, Lima, Perú.
- Hou J. (2006). A measurement of the small business Credit Gap and The use of Credit Scoring by small financial. (Doctor of Philosophy). University of Kentucky, Lexington, Kentucky, United States.
- Kleinbaum, D., & Klein, M. (2002). *Logistic Regression - Second Edition* – 2002, 1994 Springer-Verlag New York, Inc.
- KIM, J. (2005). A Credit Risk Model for Agricultural Loan Portfolios under the New Basel Capital Accord. Dissertation submitted to the Office of Graduate Studies of Texas A & M University.
- Kleimeier, S., & Dinh, T. (2007). A Credit Scoring Model for Vietnam's Retail Banking Market. *International Review of Financial Analysis*, 16 (5), 471-495.
- Kulkosky, E. (1996). Credit Scoring Could Have a Downside, Experts Say. *American Banker*, 161(208), 8.

- Limsombunchai, V., Gand, C. & Lee, M. (2005). An Analysis of Credit Scoring for Agricultural Loans in Thailand. American University of Sharjah, Sharjah United Arab Emirates
- Makowski, P. (1985).Credit Scoring BranchesOut:Decision Tree-Recent Tech-nology. Credit World, 75, 30-37.
- Mester, I. (1997). What's the Point of Credit Scoring? Business Review (Federal Reserve Bank of Philadelphia), September/October, 3-16.
- Milena, E., Miller, M., & Simbaqueba, L, (2005). The Case for Information Sharing by Microfinance Institutions: Empirical Evidence of the Value of Credit Bureau-Type Data in the Nicaraguan Microfinance Sector . New York: The World Bank, mimeo.
- Miller, M., & Rojas, D. (2005). Improving Access to Credit for Smes: in Brazil, Colombia, & Mexico. New York: The World Bank.
- Mures, M, J., Garcia, A., & Vallejo, M, E. (2005) Aplicación del análisis discriminante y regresión logística en el estudio de la morosidad de las entidades financieras. Comparacion de resultados , Pecunia, 1 , 175-199.
- Naeem Siddiqi Credit Risk Scorecards Developing and Implementing Intelligent Credit Scoring. Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
- Orgler, Y, E. (1971). Evaluation of Bank Consumer Loans with Credit Scoring Models. Journal of Bank Research, 2, 31-37.
- Orgler, Y. E. (1970). A Credit Scoring Model for Comercial Loans.Journal of Money, Credit and Banking, 2(24), 435-445.
- Puertas, R. & Marti, M. L. (2012). Análisis Del Credit Scoring. Revista Administración de Empresas, 53(3), 303–315.

- Rayo Cantón, S. Diseño de un sistema de Credit Scoring para entidades de microfinanzas. Superintendencia de banca y seguros SBS – Banco Interamericano de Desarrollo. Recuperado de http://www.sbs.gob.pe/Portals/0/jer/pres_doc_basilea/I_MODELO_DE_CREDIT_SCORING_EN_IMFs.pdf
- Rayo Cantón, S., Lara Rubio, J. & Camino Blasco, D. (Junio, 2010). Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II. Recuperado de <http://www.scielo.org.pe/pdf/jefas/v15n28/a05v15n28.pdf>
- Reinke, J. (1998). How to Lend Like Mad and Make a Profit: A Micro-Credit Paradigm Versus the Star-up Fund in South Africa. *Journal of Development Studies*, 34(3), 44-61.
- Rodríguez Guevara, D., Becerra Arévalo, J. & Cardona Valencia, D. (2017). Modelos y metodologías de Credit score para personas naturales: una revisión literaria. *Revista CEA*, 3(5), 13-28.
- Ryo H. & Hideaki h. (2010). Small business Credit Scoring: Evidence Japan. (Hosei University and Japan Center for Economic Research). Hosei University, Tokyo, Japan. Recuperado de <https://www.rieti.go.jp/jp/publications/dp/10e029.pdf>
- Saavedra-García, M. L. & Saavedra-García, M. J. (2010). Modelos para medir el riesgo de crédito de la banca *. *Cuadernos de Administración*, 23(40), 295–319.
- SBS (2008). Resolución N° 11356. Recuperado de <http://www.sbs.gob.pe/>
- SBS (2009). Resolución N° 14353. Recuperado de <http://www.sbs.gob.pe/>
- SBS (2012). Gestión avanzada de riesgo de crédito. Seminario para gerentes de riesgos de las entidades de microfinanzas. Recuperado de http://www.sbs.gob.pe/Portals/0/jer/pres_doc_basilea/I%20CONCEPTOS.pdf
- Schreiner, M.(1999). A Scoring model of the risk of costly arrears at a microfinance lender in Bolivia. Microfinance Risk Management and Center for Social Development, Washington University in St. Luis. Recuperado de <<http://info.worldbank.org/etools/docs/library/128753/Scoring%20Model%20Costly%20Arrears%20Bolivia.pdf>>.

- Schreiner, M.(2002). Scoring: The Next Breakthrough in Microfinance, Occasional Paper N° 7, Consultative Group to Assist the Poorest, Washington D.C.
- Schreiner (2001). Credit Scoring for Microfinance. Can it Work? Recuperado de <https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1098&context=esr>
- Scott Frame, W. , Padhi, M. , Woosley, L. (2004). Credit Scoring and the availability of small business Credit in low – and moderate income areas. The financial Review. Volumen (39): páginas 35-54.
- Trucharte, C., & Marcelo, A (2001). Modelos factoriales de riesgo de crédito: El modelo de Basilea II y Implicaciones, Actualidad Financiera,1,205-218
- Turvey, C.G. and R. Brown, 1990. Credit scores for a federal lending institution: The case of Canada's farm Credit corporation. Agril. Finance Rev., 50: 47-57.
- Valencia Echevarri, A.(2017). Modelo Scoring para el Otorgamiento de crédito de las Pymes. (Tesis de maestría, Universidad EAFIT, Medellin, Colombia). Recuperada de <https://core.ac.uk/download/pdf/159381488.pdf>
- Vigano, L. (1993). A Credit-Scoring Model for Development Banks: An African Case Study, Savings and Development, 17(4), 441-482.